Intrametropolitan Firm Clustering: Measurement, Detection and Determinants —A Case Study in Boston

By

Xiongjiu Liao

M.S.
University of Maryland, College Park, Maryland (2004)

M.E.
Tongji University, Shanghai, China (2002)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Urban and Regional Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

Signature of Author_____________________

Department of Urban Studies and Planning
July 18, 2011

Certified by ______________________________

Professor Joeseph Ferreira
Department of Urban Studies and Planning
Dissertation Supervisor

Accepted by____________________________

Professor Karen R. Polenske
Chair, PhD Committee
Department of Urban Studies and Planning

ACKNOWLEDGEMENTS

I would like to use this opportunity to express my gratitude for a great number of people who have contributed to my study and thesis writing in various ways. Without their help, this thesis would not have been accomplished so smoothly. First of all I would like to gratefully and sincerely thank Dr. Joeseph Ferreira for his mentorship during my stay at MIT. I appreciate his helpful comments on my thesis, which often enlightened my way, but even more so, I appreciate his willingness and availability to talk with me about my thesis at any time when I need his help. This can be a half hour during his walk to lunch or one hour at midnight when he was in another county. My experiences and leanings from his research projects helped me successfully get my first job as a project manager. I owe him a big one for all of this and I really miss the party time spent on his sail boat in the Atlantic.

I would also like to thank Dr. Karen Polenske for her guidance and encouragement on my study and thesis writing. She introduced me to regional economic modeling knowledge, which becomes one of the major foundations for this thesis study. She provided me helpful comments on thesis writing, and picked out many typos hiding in the corner.

I would like to thank Dr. Qing Shen for his continuous and generous support of my study. Dr. Qing Shen was my advisor at University of Maryland, where I accomplished my master degree in USA with his project funding. He trained me how to compute and model land use patterns and his teaching has big impacts on my further researches including this thesis study. He deserves a big thank you in all aspects from me.

Many thanks as well for Dr. William Wheaton for his teaching of urban economics knowledge and helps on my thesis at the early stage. His offering of funding to get data for this thesis study is very much appreciated and his comments on the thesis often pointed out the right working directions.

I need also thank Dr. Carlo Ratti. My participating in his innovative projects enables me to get in touch with many cutting-edge ideas and genius researchers in his group.

Sandra Wellford also deserves a big thank you from me. She helped me dealing with my status of working and studying at the same time, arranging my dissertation defense and other small but crucial procedural issues.

Furthermore, I need thank MIT library and InfoUSA providing the ReferenceUSA dataset for MIT campus users. This dataset provides detailed information about firms at establishment level, which helps our empirical studies of manufacturing firm clustering at metropolitan Boston.

Finally, and most importantly, I would like to thank my wife Lingfang Ma. She is a great lady and her support, encouragement and unwavering love had accompanied me through many difficult times during my study and thesis writing. She and our beloved children: Jason Liao and Timothy Liao were the major spiritual force supporting me through accomplishing this thesis study.

ABSTRACT

In this thesis, I attempt to construct a conceptual and computational framework for studying firm clustering at intrametropolitan level. Specifically, this framework includes methods of measuring general industry clustering, detecting sized urbanization clusters and localization clusters, interrelating firm clustering with firms' characteristics and interpreting the appearance of firm clusters at a location with its spatial settings. I apply these methods with firm establishment data to study the patterns and determinants of firm clustering within metropolitan Boston. At the first time, these studies enable us to locate urbanization clusters and localization clusters in a metropolitan area and examine their association with economic and spatial explanting factors, which provides a completely new perspective of observing those spatial economical phenomenons like agglomeration and suburbanization and makes a solid step towards empirical modeling and understanding of them at the micro-level. As demonstrated in this thesis, clustering of firms or employments both end up with significant concentration of employment at some spots with specific attributes. This enables regional planners and policy-makers to apply some findings from this study in improving existing firm clustering patterns or incubating new firm concentrations, which ultimately promote local employment and improve local economics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1. Introduction

Firm clustering, which usually refers to the fact that firms tend to concentrate in some high density areas for production or pecuniary advantages, was first noted by Alfred Marshall (1920) as "localized industries". In his book, "Book IV The Agents of Production. Land, Labour, Capital and Organization", Alfred Marshall stated that (in economic centers) *"if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the sources of further new ideas. And presently subsidiary trades grow up in the neighborhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material (p.225)*". Bertil Ohlin (1933) later developed Marshal's ideas into two "externalities" related with firm clustering: "localization economies" and "urbanization economies", where "urbanization economies" arise from the size of local economy and "localization economies" arise from the size of local industry. Today these two economic externalities are widely referenced as "agglomeration economies".

The recent boom of research efforts devoted to studying firm clustering has not been observed until 1990s (Sabel, 1989,p.18; Porter,1998, p.78; Storper,1993,p.434; Amin & Thrift,1994,p.13; Scott,1998,p.398), after the notorious success of Silicon Valley and Hollywood were widely witnessed. Since then, extensive literature has been done in studying how and why spatial proximity between firms helps promote individual firm's competitiveness and performance, in hopes of finding out the key to duplicate Silicon

Valley's myth (Neil, Eberts and Fogarty, 1993). Among these studies, two leading theories are often referenced: firm location theory, which concentrates on the individual firm's site selection on minimizing production and transport costs, and new economic geography theory, which applies general equilibrium models to show "the aggregate (clustering) behavior from individual maximization" (Krugman, 1998).

While studies that stem from firm location theories are criticized for not considering the interactions between plants (Revelle & Laporte, 1996) and largely ignoring citywide patterns (Logan, 1966), those new economic geography studies also faces critics about their oversimplification of space and neglect of social, institutional and cultural factors involved in regional agglomeration (Martin, 1999). Indeed, empirical studies at an intermediate scale between these two will have an ability to balance spatial details and generalization. This calls for spatial clustering studies at intrametropolitan level.

Intrametropolitan level studies can add knowledge of spatial clustering on three aspects, which are also the shortcomings of existing studies. First, while analysts have devoted considerable efforts in documenting the association between agglomeration and firm clustering, very few have attempted to define firm clusters clearly in terms of locations and extents. Analysts often treat firm as an elusive concept, which makes it hard to differentiate different firm clusters and even harder to answer such critical questions as the differences between firm performance inside and outside firm clusters (Appold, 1995; Swann et al., 1998). Compared with those regional level studies, which confine the models on a wonderland of no dimensions (Isard, 1949) or treat space as points (Struyk &

James, 1975, p.1), and those site studies, which often omit the existence of other economic players, intrametropolitan level studies provide a good window to identify the locations and extents of firm clusters.

Second, while regional analysts are often criticized for their biased selection of high-tech industries and regional success stories (Wiig & Wood, 1995), intrametropolitan level studies work on what Lundquist and Olander (1998) refer to as the "grey mass" of more mundane, not so spectacular firm clusters: the firm clusters around the street corners, filling in the cities and entailing a clear image of busy traffic, firms and people. In intrametropolitan level studies, after analyst identify firm clusters within metropolitan area as relationships between clustering firms and their surrounding environmental factors, such as land, transportation, policy and population become apparent and testable, which provides more representative findings than regional level studies.

Finally, while current regional analysts presuppose the existence of agglomeration economies as the only determinant of firm clustering, intrametropolitan level analysts examine comprehensive factors that involve with firm clustering, including firm characteristics, land availability and regulation, development policy, tax incitement, etc. At an age that firm clusters are more and more regarded as a development tool (Peroux, 1988, p.49; Porter, 1998, p78), an intrametropolitan level study can provide more practical findings for urban planners, economists and municipal governments to help them face competition in land utilizations. In this sense, an intrametropolitan level study will no doubt give more practical policy implications.

Although intrametropolitan level studies can provide useful findings, very few analysts have devoted time to firm clustering studies at this level, especially empirical studies, because of the limited disaggregated firm establishment datasets and the lack of modeling methods for researchers (Wu, 1999). This becomes the objective and contribution of this study. By systematically identifying firm clusters within metropolitan area and evaluating the effects of a wide range of involving factors on firm clustering, we fill a major gap in firm clustering studies at intrametropolitan level and provide effective methodological solutions for examining relationships between firm clusters and involving factors.

1.2. Study Goals, Objectives and Research Questions

Three basic questions motivating the researcher are: Which firms tend to cluster? Where do firms that cluster tend to do so? Why are these locations chosen? Answering these questions implies a need to develop an observation of spatial firm clustering at a more detailed spatial level than most of existing studies. To develop quantitative measurements of firm clustering within metropolitan areas, this study first summarizes available agglomeration indices of two groups: spatial indices and non-spatial indices, and shows what can be learned from these indices and what cannot. Then this study investigates the application of two spatial modeling methods: Geographic Analysis Machine by Openshaw (1987) and The Likelihood Ratio Test (LRT) by Kulldorff (1994) in detecting the significant firm clusters, which are rarely possible to be caused by spatial randomness even controlling for an uneven distribution of economic activities. Finally, a discrete choice model is used to examine the impacts of a list of suggested determinants by previous studies on firms' clustering behaviors.

A case study of the spatial firm clustering using industrial establishment data in metropolitan Boston provides an opportunity to test the suggested methods by the study. After measuring the overall clustering and identifying the significant firm clusters, we can model firms' decision about whether to cluster and where to cluster by using logistic regressions. We examine the importance of two groups of explanatory variables describing firm's characteristics and firm clusters' spatial attributes by using the regression.

As an aggregated result from individual firm's location decisions, firm clustering studies require an appropriate study scale. Intrametropolitan level studies provide a good balance of details versus generalization. In this study, we demonstrate that by connecting spatial measurements of the patterns of firm clustering with those detailed firm and locational attribute data, an improved understanding of firm clustering can be reached.

Two complementary goals, five research objectives and corresponding research questions provide an overall guidance and direction for this study. Figure 1-1 lists the study's goals, objectives, and research questions.

Study Goals, Objectives, and Research Questions

Study Goals:

- *G1*: Detect significant firm clusters within metropolitan areas
- *G2*: Evaluate the determinants of firm clustering and firm clusters

Study Objectives

- *O1*: Examine exiting agglomeration indices, evaluate their strengths as well as limitations
- *O2*: Develop methods to detect significant firm clusters within metropolitan areas
- *O3*: Develop methods to detect significant firm clusters within metropolitan areas with the unevenly distributed background population being controlled for
- *O4*: Develop statistical models to evaluate the impacts of explanatory factors on firm clustering
- *O5*: Develop statistical models to evaluate the impacts of locational attributes on location of clusters

Research Questions:

- *RQ1*: Are industry firms clustering within metropolitan Boston? If so, at which scales?
- *RQ2*: Where do industry firms tend to cluster within metropolitan Boston?
- *RQ3*: Where do industry firms tend to cluster when unevenly distributed background population is controlled for?
- *RQ4*: Which firms tend to locate within clusters?
- *RQ5*: Where do firms that tend to cluster do so?

Figure 1-1: Study Goals, Objectives and Research Questions

The first study goal (G1) is to detect the significant firm clusters within cities. Chapter 3 - 5 introduce three topics about this goal: how to measure overall clustering, how to locate significant firm clusters, how to locate significant firm clusters while controlling for uneven distribution of economy.

The second study goal (G2) is to evaluate the determinants that affect firms' clustering behaviors about whether to cluster and where to cluster. Chapter 6 introduces the logistic regression method to test the impacts of firm characteristics on firm clustering. In chapter

7, we apply the spatial auto-logit to evaluate the spatial attributes' influence on locations of firm clusters.

1.3. Contributions of Study

The purpose of this study is to fill the gap of firm clustering studies at intrametropolitan level by offering an empirical examination of the patterns and determinants of firm clustering within Boston metropolitan area. This study contributes to the literature of firm clustering and agglomeration in four fields. First, a review and analysis of related literatures about firm clustering and agglomeration is conducted. Despite the fact that a large volume of literatures have been found, this review leads to conclude that there is very limited work on firm clustering at intrametropolitan level. In particular, this review finds that there have been no studies devoted to examining the firm clustering patterns within metropolitan area systematically. Second, this study empirically analyzes the firm clustering patterns within metropolitan Boston and presents applicable spatial modeling techniques for detecting the significant clusters. Third, this paper steps further by empirically investigating the determinants of firms' firm clustering and firm clusters, which are mostly left unanswered in previous studies. Finally, this study provides tractable methods for computing firm clustering using the most disaggregated firm datasets available within metropolitan areas.

1.4. Study Methodology

The study is designed to be an empirical and quantitative research. The study focuses on developing and applying spatial and statistical models to measure and understand firm clustering within metropolitan areas using detailed firm establishment datasets. The study employs a case study that provides a methodological structure for testing the suggested methods in measuring the overall clustering, detecting significant firm clusters and evaluating firm clustering determinants.

At the outset of the study, the research introduces the overall study methodological framework for the study. From Chapter 2 to Chapter 7, specific methodologies are introduced at the beginning of each chapter when answering corresponding research questions. This approach avoids the separation of introductions and discussions of methodologies from their research question.

The study applies two spatial models: Geographic Analysis Machine by Openshaw (1987) and The Likelihood Ratio Test (LRT) by Kulldorff (1994) to capture the statistical features of firm establishments' spatial distribution. Monte Carlo simulation is used to test the statistical significance of the results from these models. First introduced by Hope (1968), Monte Carlo significance test is accomplished by comparing the observed test statistic with random samples that are generated in accordance with the hypothesis being tested, for example complete spatial randomness (CSR) or unevenly distributed population data. These spatial analysis procedures involve intensive data calculation. Geographic information system (GIS), specifically, ESRI's ArcView 3.1 is used to assist this work.

Logistic regression model is used to model firms' clustering behaviors by observing firms' decisions over whether to cluster and where to cluster based on firms' independent characteristics and locations' unvaried attributes. It does not mean that firms only consider these two issues during their location process. Other decision considerations are omitted because they are irrelative to our discussion. Chapter 6 and 7 provide detailed discussions about the models we applied.

1.5. Structure of the Study

The rest of study is organized as follows. In chapter 2, we review the literatures from three relevant fields: firm location theory, new economic geography theory and agglomeration and discuss their references for understanding intrametropolitan firm clustering. We then survey related empirical studies of firm clustering and identify the practical and theoretical values of conducting firm clustering study at intrametropolitan level.

In chapter 3, we begin to answer the question of whether there exists firm clustering within metropolitan Boston by measuring its overall firm clustering index. We choose two widely referenced measurements: concentration index (Ellison & Glaser, 1997) and the distance-based index (Marcon & Puech, 2003) to measure the aggregated firm clustering index within metropolitan Boston. Findings from these two methods are compared to show how different aspects of firm clustering can be demonstrated using a non-spatial index like concentration index and a spatial index like the distance-based index. The limitations of these two methods, such as Modifiable Areal Unit Problems (MAUP), showing the location of clusters, are also discussed.

In chapter 4, we go further to discuss another two important characteristics of intrametropolitan firm clustering: the location and extent of firm clusters. We demonstrate how to apply Geographic Analysis Machine (Openshaw, 1987) to scan a metropolitan area for detecting significant firm clusters at various spatial scales, as well as how to test the statistical significance of the measurements.

In chapter 5, we expand the discussions in chapter 4 to detect those clusters specifically favored by manufacturing firms. Because downtown areas are usually identified as the dominant firm clusters in terms of absolute firm number, it is of interest to identify and analyze those "hot spots" of manufacturing firms. In this chapter, we present how to apply The Likelihood Ratio Test (LRT) (Kulldorff ,1994) to map the manufacturing firm clusters while controlling for unevenly distribution of economy. Statistical reference strategies about these measurements are also discussed.

After mapping firm clusters in chapter 4 and 5, in chapter 6, we construct logistic regression model to examine the impacts of firms' characteristics on their decisions over whether to cluster. In chapter 7, a spatial autologit model is adopted to study the impacts of locational variables on firm clustering. The results provide a reference for answering two core questions about firm clustering at intrametropolitan level: Which types of firms tend to cluster? Where do firms that tend to cluster do so?

In chapter 8 we conclude our findings from the study and list further study directions.

CHAPTER 2: LITERATURE REVIEW

2.1. Review of Existing Literature

The explanation of firm clustering can be traced back to Alfred Marshal's (1925) statement about "technical spillover", which suggests that in economic centers knowledge sharing among people will spur further new ideas thus promote economic growth. After Marshal, Alfred Weber further points out firms tend to locate close to economic centers to take advantage of the sufficient labor, demand and services there(1929). In Weber's model, the appearance and interrelations of involving forces come to play in historical sequences when people occupy an undeveloped country and establish an isolated economic system (Isard, 1949). This static way of dealing with land related factors was criticized by later researchers and they proposed a general equilibrium approach, which solves inputs caused by distance and land simultaneously with labor and capital inputs. (Predöhl, 1928; Englander, 1926; Isard, 1949). In these studies, Weber's agglomeration effects were explained as the potentialities for substitution between various types of costs, chiefly between land cost and costs on labor, capital or transportation or the combination of these three.

A major problem with these general equilibrium efforts is their oversimplifying space by treating locations as points, whose only attribute is the transportation distance to suppliers or clients, as stated by Predöhl (1928,p.371-390), "*if there is any sense at all to location economics, it is because there are certain regularities in the variations of costs and prices over space. These regularities arise primarily because transport cost is some* *function of distance.*" This statement is surely too narrow in the sense that it neglects all other social-economic details about space.

After a period of relatively neglect in 1970s and 1980s, in 1990's much literature of firm clustering were observed in response to the notorious success of some firm clusters, four most celebrated examples of which are Silicon Valley's semiconductor industry, Hollywood's movie business, and the pharmaceutical industry in New Jersey and Massachusetts. Extensive studies have been devoted to understanding the role of firm clustering in promoting firm competitiveness and regional economic growth. Among them, a group of researches have generated the greatest attention under the name of "new economic geography", which shortly becomes "one of the most exciting areas of contemporary economics" (Ottavino & Puga 1997; Schmutzler, 1999). The representative works in new economic geography theory include Krugman (1991, 1995, 1998), Krugman and Venables (1995), Venables (1996), Fujita and Krugman (1995), and Puga and Venables (1996, 1997).

In new economic geography theory, clustering of firms is explained with the existence of pecuniary externalities, which arise from interactions of increasing returns with transportation costs between regions. To illustrate how the pecuniary externalities generate spatial firm clustering, here we use an example given by Meardon (2001): Suppose there are two regions A and B initially with an equal number of firms, labor and other resources. One day region A, for some reason, gains one more firm. Now region A becomes more agglomerated than region B and the firms in A enjoy larger pecuniary externalities than those in region B. Moreover, in region A, the consumer price index decreases as consumers save transportation cost when producers concentrate together and the nominal wages increase as the productions becomes greater, which increases the labor demands. The lower price and higher wage further attract labor to migrate from region B to region A. When the incoming of new labors decreases the nominal wages in region A, it also creates another profit opportunities to additional firm.

While it is said that the formal-model-driven nature of new economic geography theory helped it quickly to be accepted by mainstream economists (Meardon, 2001), it is also from where the major critiques come. The critics over new economic geography are almost identical as those critics about old industrial location theories: oversimplification of space and exclusion of those spatial details, about which Ron Martin (1999) gave a good summary:

> "*Now, clearly, there are aspects of economic development in general, and spatial agglomeration, in particular, that do lend themselves to mathematical representation and modeling. But there are also severe epistemological and ontological limits to such a narrow approach. For one thing, it means that "messy" social, cultural, and institutional factors involved in spatial economic development are neglected. Since these factors cannot be reduced to or expressed in mathematical form they are assumed to be of secondary or marginal importance and, as Krugman puts it, are 'Best left to sociologists.' But it is precisely the social, institutional, cultural and political embeddedness of local*

and regional economies that can play a key role in determining the possibilities for or constraints on development, and thus why spatial agglomeration of economic activity occurs in particular places and not others (Martin, 1999; p.75)."

Without detailed treatment of locations, new economic geography models are criticized for only giving general conclusions about firm clustering but telling no details about it: where, why and how firm clustering happens in a specific region (Martin, 1999, p.389).

Compared with the oversimplification of firm clustering in theoretical studies, empirical studies have discovered a variety of factors that can affect firm clustering, for example: labor market pooling (Helsley & Strange, 1990), inter-medium product input (Goldstein & Gronberg, 1984) and knowledge spillovers (Glaeser, 1999). Krugman (1980) also finds that firm clustering at specific locations can be caused by historical accidents and once the clusters are established, they get "locked in" at these accidental locations because of cumulative gains from trade. Saxenian (1994) argues that a region's performance in industry networking and supporting start-ups will affect its firm clustering and two often referenced examples of this factor are Silicon Valley's continuing to be clustering and the declining of Route 128 corridor in Massachusetts.

Wolf and Pett (2000) find that firm size also plays an important role in firm's clustering. As large firms have the capacity to internalize many of the advantages of clustering (labor talent, access to markets, business services), they are more reluctant to locate within firm clusters. Small firms, on the contrary, are willing to endure the diseconomies of urbanization to gain otherwise unavailable external scale economies by locating within firm clusters. Phelp, Fallon and Williams (2001) discover that new firms try to locate in close spatial proximity of the existing clusters in order to utilize economic externalities. However, not all findings are consistent. Honjo (2000) observes firm clustering causes negative impacts on the survival of new firms because of the resulted occurrence of higher competition in the center and Okamuro's (2008) recent research shows that larger firms' uncooperative attitudes can hinder small firms' enthusiasm in locating within clusters.

Researchers also find firm's organizational structure can impose impacts on firm's clustering behavior. DeYoung, Robert et al. (2004) observes that while bank headquarters move to those larger cities with agglomeration economies available to banking companies, bank branches have moved substantially farther away from bank headquarters. Hong (2007) also finds that compared with independent firms, branch companies are less responsive to local market demand and labor factors, since for branch companies, scale economies and knowledge transmission is mostly internalized within the company.

Studies have shown that firm clustering exists at various spatial scales, from street corners, metropolitan areas to a nation. Marcon and Puech find that firm clustering varies continuously with the change of study scales (2003). Rosenthal and Strange (2001) point out that the effects of firm clustering determinants vary with study scales. That is to say, clustering determinants exhibit different impacts on firm clustering with the change of spatial scales and findings at large scales cannot always hold at smaller scales.

Instead of just being recipients of firms, space's site attributes are also found to influence the appearance of firm clustering on specific location. In his study of the location choice of manufacturing plants in the State of Sao Paulo, Brazil, Hansen (1987) discovers that local agglomeration level is significantly associated with firms' concentration and a location's proximity to city centers is also an important consideration for firm location.

In the literature, we also observe a trend of applying firm clustering as a regional economic development policy, based on the recognition of the imperfect competition and increasing returns within the clusters (Martin & Sunley, 1996). Cluster policies have been advocated to stimulate competition while concentrating firms in the same industry, help diversify local economic base and better withstand downturns in the economic cycle (Leveen, 1998; Doeringer & Terkla, 1995). Porter (1997) specifically proposes to create an industry cluster within inner city since some industries will gain competitive advantage by locating there. However, not all researchers are proponents for cluster policies. Harrison and Glasmeier (1997) and Rosenfeld (1995; 1997) criticize industry clusters' reluctance to large technology and market changes and the potential crisis while over-specializing in the local economy.

2.2. Limits and Challenges in Existing Research

Through literature review, we observed three important limits in existing studies about firm clustering. First, although analysts verify that agglomeration economies attenuate quickly within a few miles, which clearly implies the existence of boundaries around firm clusters, few of the studies have delineated the boundaries of firm clusters. The closest attempt is the work done by Marcon and Puech (2003) in measuring the spatial concentration of firms within a metropolitan area. However, they only show the degree of aggregated firm clustering varying at different spatial scales and the locations and boundaries of firm clusters are left unidentified. In existing literature, those critical questions related with particular firm clusters are left unanswered: the differences in firm performance between those located inside or outside firm clusters (Appold, 1995; Swann et al., 1998), the precise mechanisms behind and the magnitude of localization economies within firm clusters (cf. Sabel, 1989; Malmberg ,1997;Larsson, 1998).

Second, while literature shows that firm clustering appears at different spatial scales, very few researchers study the distribution of firm clusters within a region. Do firm clusters overlap? Are firm clusters of similar sizes? Do firm clusters at different spatial scales exhibit similar distribution pattern? What are the interactions among firm clusters? Unfortunately, we can get very limited information about these questions from existing literature.

Third, analysts have shown that firm characteristics affect firm's clustering behaviors and locational attributes also affect the local milieu for firm clustering. However, few analysts have systematically evaluated the magnitudes and directions of their impacts. Bull (1985) gives the closest attempt in his industrial location study, where he finds that a firm's intra-urban locations can be better understood by the site-specific attributes. In the study of Bogota, Lee(1989) differentiates sites with a variety of attributes, including proximity to markets and suppliers, distance to the labor market, availability of utilities and municipal services, etc. However, these studies are not linked to a firm's clustering behaviors.

2.3. Summary

In summary, existing efforts on firm clustering, especially those at regional studies, have run into difficulties when trying to identify the precise magnitude and mechanism of firm clustering within particular clusters. Empirical studies have to focus on delineated firm clusters that are differentiated from other locations by a variety of variables describing firm clustering. To get concrete progress in firm clustering studies, the extends and locations of firm clusters ought to be determined at intra-metropolitan (or intra-urban) level to define clearly the study objects: the clustering firms or the clustering sites. In this way, the differences of firm characteristics, productivities and transactions between firms inside clusters and outside clusters can be identified and compared to understand the clustering mechanisms. Similarly, analysts can study the differences of locational attributes between sites within clusters and outside clusters to understand the locational preference of firm clusters.

This review leads to the conclusion that there exits little work on firm clustering at the intrametropolitan level. In particular, there have been very limited studies devoted to examining the firm clustering patterns systematically within a metropolitan area and evaluating their determinants. This is what our study will contribute. With use of detailed firm establishment data and by developing spatial modeling techniques for detecting the significant clusters, we are able to empirically examine the patterns and determinants of firm clustering within Boston metropolitan area, which are mostly left unanswered in earlier studies.

CHAPTER 3: MEASURING INDUSTRY CLUSTERING

3.1. Introduction

There exist two tendencies when defining industry clustering: one views industry clustering as a geographic concentration of the employment or production in an industry that cannot be explained by chance alone (Ellison and Glaeser, 1997; Krugman, 1991; Rosenthal and Strange, 2001; Marcon and Puech, 2003; Porter, 1990.). The other views it as the collocation of a group of interrelated industries that are linked via supply chain or market sharing (Feser, 1998; Porter, 1997).

The deviation in definition has led to diverse methods applied in empirical studies to measure the degree of clustering. In this chapter, we follow the former interpretation of firm clustering and measure the clustering of manufacturing industry within Boston metropolitan area.

3.2. Review Industry Clustering Measurement Methods

A large number of methods have been proposed and published for measuring industry clustering, which can be coarsely divided into two categories according to whether they are location dependent: spatial index and non-spatial index. The former includes Besag's L(r) function and Diggle and Chetwynd's D(r) function. The latter includes such indices as location quotient, herfindahl index, gini index and EG concentration index. Going through these methods represents a demanding task, but it provides a big picture of existing efforts in understanding firm clustering and identifies the direction and significance of further studies in the following chapters.

3.2.1 Non-spatial Indices

a) Location quotient

Location quotient is one of the most widely used indices analysts use for studying industry clustering, and they have widely used it in economic geography and regional economics since the 1940s (Gibson, Miller and Wright, 1991: 65). This technique has remained popular in large part because it requires little data and analytical skill and can be carried out quickly and inexpensively (Isserman, 1977, p. 33). A location quotient (LQ) is calculated as: $LQ_i = (e_i / e)/(E_i / E)$,where: e_i is the employment of specific industry within the study area, e is the total employment in the study area, E_i is employment of specific industry in the reference area, and E is total employment in the reference area. When the LQ is less than 1.0, it means that the study area has a lower ratio of employment in the specified industry relative to the reference area and Vice Versa. Generally speaking, the higher the location quotient is, the more clustering of the specified industry in the study area, compared to that of the reference area.

The major limit of the location quotient is its solely dependency on data about the specified industry without relating with the population size, industry size and industry spatial distribution. The results of LQ analyses are often criticized to be too simple to give any meaningful information about industry clustering.

b) Herfindahl index

The Herfindahl index, also known as the Herfindahl-Hirschman Index, or HHI, is a widely used measure of industry concentration. It is calculated as the sum of the squared market share of each firm in the market. For example, for an industry with five firms with shares of twenty, thirty, ten, five, twenty and fifteen percent, its HHI is 0.165 ($0.2^2 + 0.3^2 +0.1^2 +0.05^2 + 0.15^2 = 0.165$). The two extreme cases are: (1) the industry consists of many (n) equal sized firms, where the HHI is 1/n (close to 0 if n is large enough), and (2) the industry consists of only one dominant firm, where the HHI is 1. Thus the HHI is with a value between 0 and 1.The smaller the HHI is, the more equally distributed the industry, and vice versa.

c) Gini coefficient

The Gini coefficient was first proposed by the Italian statistician Corrado Gini in 1912 (1912) as a summary statistics to measure inequality in a distribution. Since then, the Gini coefficient has been widely accepted by economists in studies, especially in income inequality and poverty studies (Anand, 1983; Chakravarty, 1990; Lambert, 1989; Silber, 1999; Atkinson & Bourguignon, 2000; Briant, Combes, & Lafourcade, 2007;). Before the Gini coefficient can be calculated, Lorenz curve is first drawn. Lorenz curve is a graphic curve representing the relationship between the cumulated employment (or income) share in an industry (on the vertical axis) and that of all industries (on the horizontal axis), when the data units are ranked by the ratio of these two shares. The Gini coefficient is defined as the area between the line of equality and the Lorenz curve versus that under the Lorenz curve. When the size distribution of an industry is proportional to that of the total employment in all industries in a region, it has a Gini coefficient of 1, since the Lorenz curve would be a 45-degree straight line. On the contrary, if a region has all of its employment in one industry concentrated in one firm, it will have a Gini coefficient of 0, since its Lorenz curve would just be a point on the vertical axis.

d) EG concentration index

Ellison and Glaeser (1997) proposed a concentration index (EG index) to measure the deviations of the observed concentration of an industry from the concentration that would result from firms locating randomly and independently. The EG index can be viewed as an evolvement from Gini coefficient. In their study, Ellison and Glaeser proved that the expected value of the spatial Gini coefficient (G) can be represented as:

$$E(G) = (1 - \sum x_i^{2})[\gamma + (1-\gamma)H], \qquad (3\text{-}1)$$

where γ is the index measuring the concentration due to natural advantages and agglomerative externalities among firms. From Equation (3-1), Equation (3-2) can be achieved with a straightforward deduction.

$$\gamma = \frac{G - (1 - \sum_i x_i^{2})H}{(1 - \sum_i x_i^{2})(1 - H)}, \qquad (3\text{-}2)$$

where:

G: spatial gini index, calculated as: $G = \sum_i (s_i - x_i)^2$,

H: Herfindahl index, calculated as: $H = \sum_i s_i^{2}$,
s_i: location i's share of employment in the studied industry,
x_i: loation i's share of total employment.

EG concentration index are widely applied in industry clustering studies across the world. For example, in the United States by Ellison and Glaeser (1997), Rosenthal and Strange (2001); for the United Kingdom by Devereux et al. (1999); for Spain by Callejòn (1997); for Rance by Maurel and Sedillot (1999), Houdebine (1999); etc. However, two major deficientcies of EG concentration index are found in empirical studies. First, researchers find that its value changes with geographic units used, which is often termed as "Modifiable Spatial Unit Problems" (MAUP). For example, Rosenthal and Strange (2003) find that EG indexes based county and zip code level data are much smaller than those based on state level data. Second, while it gives a general description about the degree of concentration of a region, the EG index tells nothing about the extents and locations of industry clusters.

3.2.2 Spatial Indices

a) L (r) function

Noticing that EG concentration index can only measure industrial clustering at a single scale, Marcon and Puech (2003) proposed to use Besag's L(r) function (1977) to measure the spatial concentration of firms. L function is a modification from the K function, which was first proposed by B.D.Ripley (1976,1977). Ripley defines K(r) as the average number of neighbors divided by λ, which is the average spatial density of firms within the study area, and proves that under complete spatial randomness (CSR), K(r) equals to πr^2. CSR is defined as the distribution that firms choose their locations within the study region randomly and independently. Thus by comparing the observed k(r) with πr^2, we can know whether a distribution is more concentrated or dispersed than CSR. To avoid the multiple comparing at different radius, Besag (1977) simplifies K(r) and proposes L(r) to be used. The L function is defined as follows:

$$L(r) = \sqrt{\frac{K(r)}{\pi}} - r, \qquad (3\text{-}3)$$

where,

r: the radius of the circle which centers on each of the firm establishments,

K(r): $K(r) = \frac{\bar{n}}{\lambda}$ ($\bar{n}$: average number of firms found within radius of r; λ is the average spatial density of firms within the study area and $\lambda = \frac{N}{A}$; N: total number of firms; A: the area of the study area).

With L(r), we can determine whether the distribution of firms is significantly different from CSR by comparing L(r) with 0. L(r)>0 indicates that the observed distribution is geographically concentrated, while L(r) <0 implies dispersion.

b) D(r) function

In calculation, L(r) function uses CSR as benchmark, but random distribution also causes concentration, which makes the benchmark less meaningful. For example, throwing dart to the dartboard with eyes covered, after certain repeat times, we normally observe heterogeneous distribution on the board, instead of a homogeneous one. Under this circumstance, Diggle and Chetwynd (1991) proposed D(r) for measuring the relative clustering of an industry. Instead of comparing the distribution of the case industry to CRS, D(r) compares it with a control industry. Thus D(r) indicates whether the case industry is more or less clustered than the control industry. For example, Sweeney and Feser (1998) apply D(r) to investigate whether small firms are more concentrated than large firms. Diggle and Chetwynd show that if the case distribution and the control distribution are identical, we will observe $K_{case}(r)=K_{control}(r)$ at any distance r. The difference of the two distributions: $D(r)= K_{case}(r)-K_{control}(r)$ defines a benchmark to determine whether case distribution is more concentrated ($D(r)>0$) or more dispersed ($D(r)<0$) than the control distribution.

3.3. Measuring Industry Clustering in Boston

Above indexes measure industry clustering from different perspectives and provide important information about industry clustering in certain aspects. In this section, we apply above measurements, including the Herfindahl index, Spatial Gini, EG concentration index, L(r) and D(r), to evaluate the degree of concentration in Boson's manufacturing industry. Throughout the study, we discuss the advantages and limits of these indexes in empirically measuring industry clustering at intrametropolitan level.

3.3.1 Data

To compute these indexes, we use information from the U.S. Business database of ReferenceUSA in 2009. The complete version of U.S. Business database contains information from over 14 million firm establishments in the United States. We utilize the datasets for firms within Boston metropolitan area. This represents the best disaggregate firm datasets in the United States, compared with Ellison and Glaeser's data at the county-level, Rosenthal and Strange's data at the zip level. This dataset is also better than those used in studies outside of the United States, for example Briant, Comes and Lafourcade's data for Cantons in France(corresponding to census tract in United States), Marcon and Puech's dataset at the zip code level.

Due to the limitation of available time, we have chosen to focus only on manufacturing industries as detailed as four-digit level, which is often chosen by researchers to balance location fineness and industry fineness. Using the dataset, we compute the Herfindahl index, spatial Gini index, EG concentration index at the Standard Industrial Code (SIC) 4-digit level, and calculate L(r) and D(r) at firm-establishment level.

3.3.2 Study Area

All townships within Highway 495 in Boston are chosen as the study area, which we refere to as the metropolitan Boston, as shown in Fig 3-1. This area is considered not only because it is the place that attracts traditional academic interests, but it is the place with the most active industry clustering phenomenon in New England region. Within the study area, we observe 158301 business establishments, among which 7828 (or 5%) are coded as manufacturing firms. Fig.3-1 shows the study region we use for Boston, and Table 3-1 gives summary information about our database.

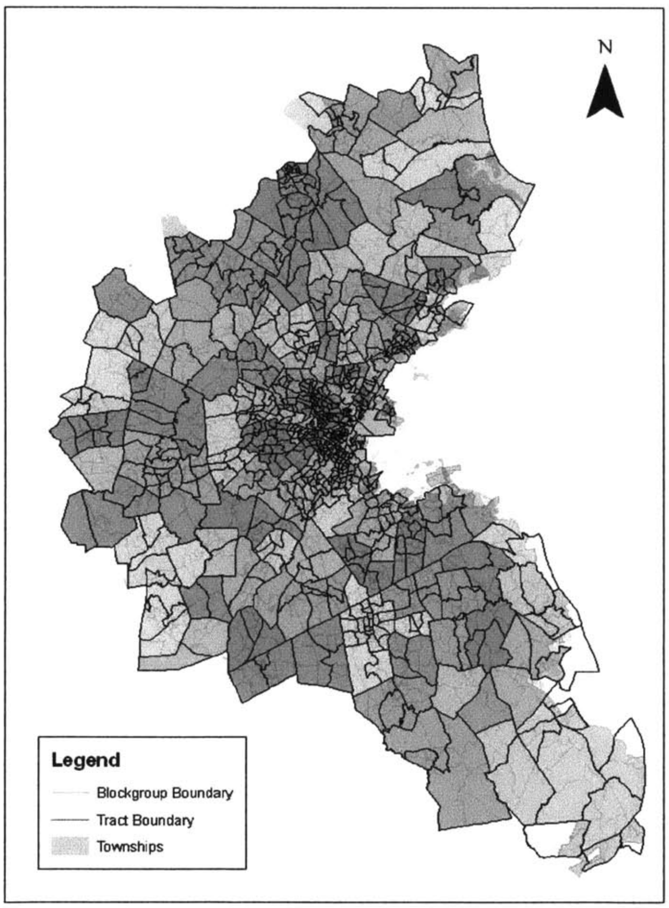


Figure 3-1 Boston Metropolitan Area

3.3.3 Computation and Related Issues

a) *Herfindahl, spatial Gini index and EG concentration index*

The calculation of these indexes is pretty straightforward. After adding up the industry total employment (at 4-SIC digit level) from establishment location employments in this industry, we estimate the Herfinahl index by summarizing the squared share of each establishment' location employment of the total industry employment. We calculate the spatial Gini index as the sum of each sub-area's squared differences in terms of share of manufacturing employment and that of an industry employment. To calculate EG concentration index, we need also calculate each sub-area's squared share of manufacturing employment besides spatial Gini index and Herfindahl index. Then we calculate the EG concentration index as shown in equation (3-1).

To compare the concentration measurements at different spatial units, we calculate these indexes at three levels: town, census tract and block group. Fig. 3-1 shows these three levels: the shaded polygons represent 127 towns within 495 highway; the dark lines are the boundaries of 737 census tracts, and the grey ones are the boundaries of 2814 block groups. As we will discuss this issue in following sections, our findings indicates the existence of variances in the measurement results with the change of spatial unit adopted, suggesting the existence of modifiable spatial unit problems (MAUP).

b) L(r) function and D(r) function

Three steps are taken to calculation L(r) function: First, a 1km*1km grid is laid over the study region. In total there are 5579 grids produced for study area; 19 circles of radiuses ranging from 1km to 10km with 0.5km increase step are produced around the center of each grid. The number of manufacturing establishments is counted for each circle and the average of these numbers across all grids is K(r). finally, L(r) can be calculated according to equation (3-3) with λ=0.12 (manufacturing establishments)/ km^2.

Similarly, D(r) is calculated as the difference of K(r) for manufacturing sector and the K(r) for all sectors. We want to find out whether manufacturing firms are more concentrated compared with all other industries within Boston metropolitan area.

c) The problem of "Edge effects" and correction

When calculating L(r) functions and D(r) functions, we need use circular zones to count number of firm establishments within each zone. When circles are close to the boundary, the count will be underestimated because portion of the circles will be outside of the study area. This problem is called the "edge effects". Researchers have taken different strategies to correct "edge effects". For example, Openshaw and Chalton, et al. (1987) surround the study region with an internal or external corridor equal in width to the radius of the largest circle; Diggle (1983) and Rowlingson and Diggle (1993) choose to correct the counted number by a factor equal to the circle's portion of perimeter within the study region; Marcon and Puech (2003) adopt a similar strategy with Diggle and Rowlingson, but they use the area rather than the perimeter portion of circles within the study area as the ratio to correct the counted number. This is also our choice for this study. That means if n firm establishments are observed within a circle, which has p percentage area within the study area, then the number of firm establishments is adjusted to be n/p, instead of n.

d) Statistical Inference

For the Herfindahl index, spatial Gini index and EG concentration index there does not exist statistical inference. The measurement results are interpreted according to human experiences. For example, US Department of Justice regard market with Herfindahl index greater than 0.18 as concentrated. And Ellison and Glaeser regard an index highly concentrated if the EG concentration index is greater than 0.05; not very concentrated if EG is smaller than 0.02.

For the results of L(r) functions and D(r) functions, they will face significance test by using Monte Carlo significance Test procedure. Monte Carlo procedure was developed by Hope (1968) and since then it has been widely used because of its computational and conceptual simplicity. Monte Carlo significance test procedure is accomplished by comparing the observed test statistic with random samples generated in accordance with the hypothesis being tested. The outcome of the test is determined by the rank of the test statistic relative to the random samples forming the reference set. The number of random samples plus one would constitute the significance level. For example n=99 random samples will produce a 0.01 level. Lower significance level can be reached by increasing the number of random samples.

To determine the statistical significance of the results of L(r) functions, in the study we also calculate the distribution of L(r) based on random samples at the 0.01 significance level, which provide a benchmark for us to judge the significance of our results.

3.3.4 Results and Discussion

a) *Non-spatial indices*

To measure the Boston firm establishment data with three indices: Herfindahl index, spatial Gini index and EG concentration index. These three indices are widely applied in researches, however few literatures in which the three indexes are compared. Because the Herfindahl index and the spatial Gini index are used as components in calculating the EG concentration index, which has become a popular choice for measuring concentration, it is beneficial to understand their specialties when reading reports of them from different studies. Our concern of MAUP in the calculations of spatial Gini index and EG concentration index causes us to conduct the measurements at three different spatial scales: census tract (call Tract in the following text), block group (call Group in the following text) and town (call Town in the following text). Although we could not find them in previous literature, these spatial units are not the most detailed spatial scales among existing literature, since Rosenthal and Strange (2001) calculate EG concentration index with the zip code level data. However, these three spatial units provide much more details about industry clustering at local scales than previous studies, most of which focus on regional scales.

In their study, Rosenthal and Strange (2001) find that when they go from 2- to 3-SIC digit industries and from 3- to 4-SIC digit industries, the average level of concentration increases, because when industries are aggregated into broader and fewer categories, "spatial patterns of establishment locations eventually approach that of the entire economy, causing G and γ to shrink toward zero" (Note: G refers to spatial Gini index; γ refers to EG concentration index). Thus we choose to focus our study on the 4-SIC digit manufacturing industries. As the small number of firm establishments in an industry will greatly affect the value of Herfindahl and EG concentration index, to avoid the impacts of small sample problem, we select those industries with more than 10 (including 10) establishments for the study. This gives us 146 industries of original 369 industries.

- **Are manufacturing industries concentrated?**

Table 3-1 lists the summary information of the measurement results of Herfindahl index, spatial Gini index and EG concentration index. From the table, we observe obvious deviation from zero from all of the three indexes. Their mean or median value is about one standard deviation from zero values, which means about 16% possibility to be explained by data, assuming it is normally distributed with mean of zero. It is not a statistically significant number, but it does disclose an obvious concentration tendency within the data. This is more clearly shown by Table 3-2, which represents the distribution of the EG concentration index for three spatial units we chose.

From Table 3-2, most of industries have EG concentration index greater than zero. Although there does not exist significance test for EG index, $\gamma>0$ is interpreted as a signal of the existence of excess concentration. While this study verifies the ubiquitous concentration among industries, the degree of concentration is not that large. If we apply the criteria of 0.05, which Ellison and Glaeser define as the lower bound for very concentrated firm distribution, we only observe 8-15% industries are very concentrated. This is also shown by the histograms of γ at town, tract and group level (Fig. 3-2, 3-3, 3-4). The bars on these three charts represent portion of industries falling the intervals of width 0.01. From these figures, we have the same findings that significant portion of industries fall in the range of $\gamma<0.02$, which is interpreted as not very concentrated. Together, these tables and charts give an image of ubiquitous firm concentration in our data. These findings coincide with the findings by Ellison and Glaeser, Rosenthal and Strange with different datasets at different spatial levels.

Table 3-1 Measurement Results of Three Indices

	Minimum	Maximum	Mean	SD	Median	Correlation with same index at Group	Tract	Town
EG Concentration Index								
Group	-0.0749	0.2618	0.0149	0.0347	0.0137	1		
Tract	-0.0560	0.5678	0.0206	0.0578	0.0146	.992(**)	1	
Town	-0.0450	0.5770	0.0326	0.0597	0.0208	.997(**)	.990(**)	1
Spatial Gini Index								
Group	0.0234	0.8552	0.2272	0.1844	0.1737	1		
Tract	0.0269	0.8578	0.2296	0.1859	0.1779	.959(**)	1	
Town	0.0291	0.8405	0.2338	0.1885	0.1744	.989(**)	.960(**)	1
Herfindahl Index								
	0.0143	0.8608	0.2159	0.1785	0.1809			

** Correlation is significant at the 0.01 level (2-tailed).

Table 3-2 Distribution of EG Concentration Index

Range	% of Industries Group Level	Tract level	Town Level
$\gamma<=0$	16	18	10
$\gamma \in (0,0.02)$	60	48	38
$\gamma \in [0.02,0.05)$	14	26	37
$\gamma \in [0.05,1]$	8	8	15

Table 3-3 Correlation Coefficients for Three Indices

		Herfindahl	EG Concentration Index		
			Group	Tract	Town
Herfindahl		1	-.126(*)	-0.137(*)	-0.099
Spatial Gini Index	Group	0.959(**)	-0.096	-0.125(*)	-0.084
	Tract	0.943(**)	-0.086	-0.114(*)	-0.073
	Town	0.968(**)	-0.103	-0.122(*)	-0.084

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

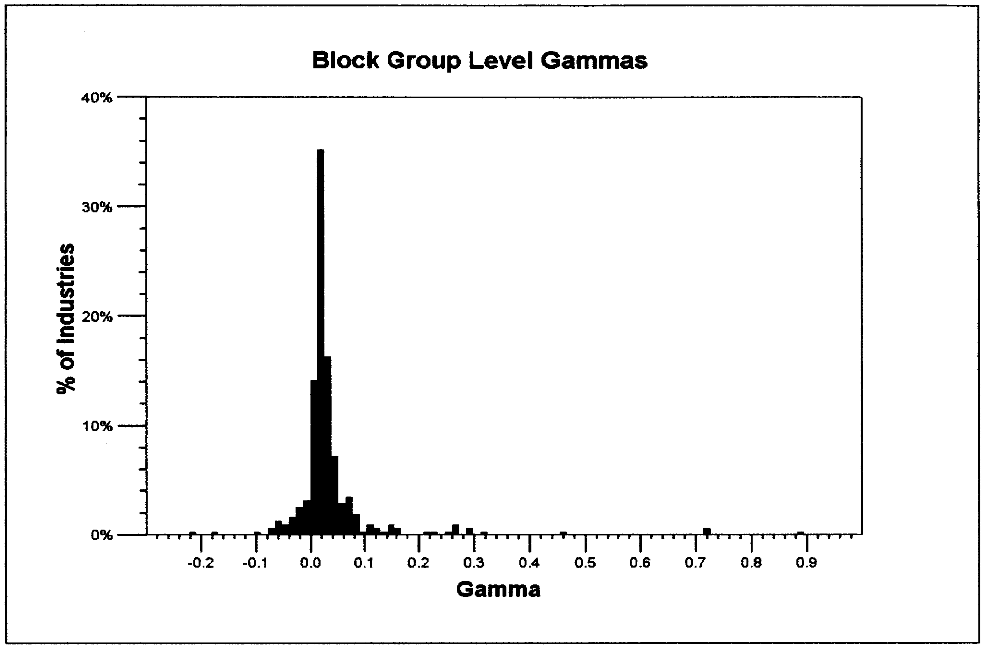


Figure 3-2 Histogram of γ at Group Level

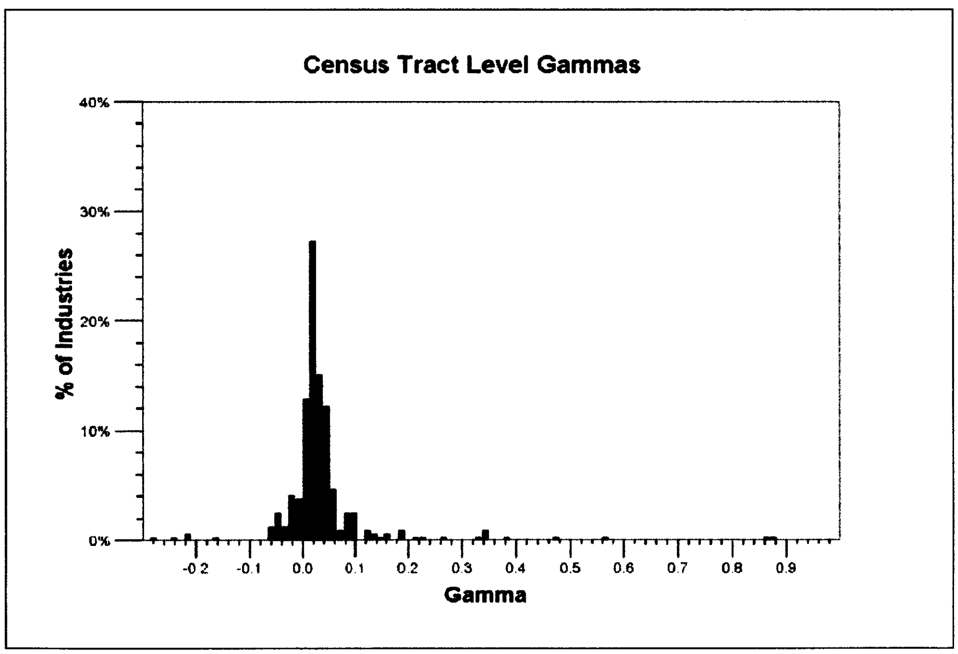


Figure 3-3 Histogram of γ at Tract Level

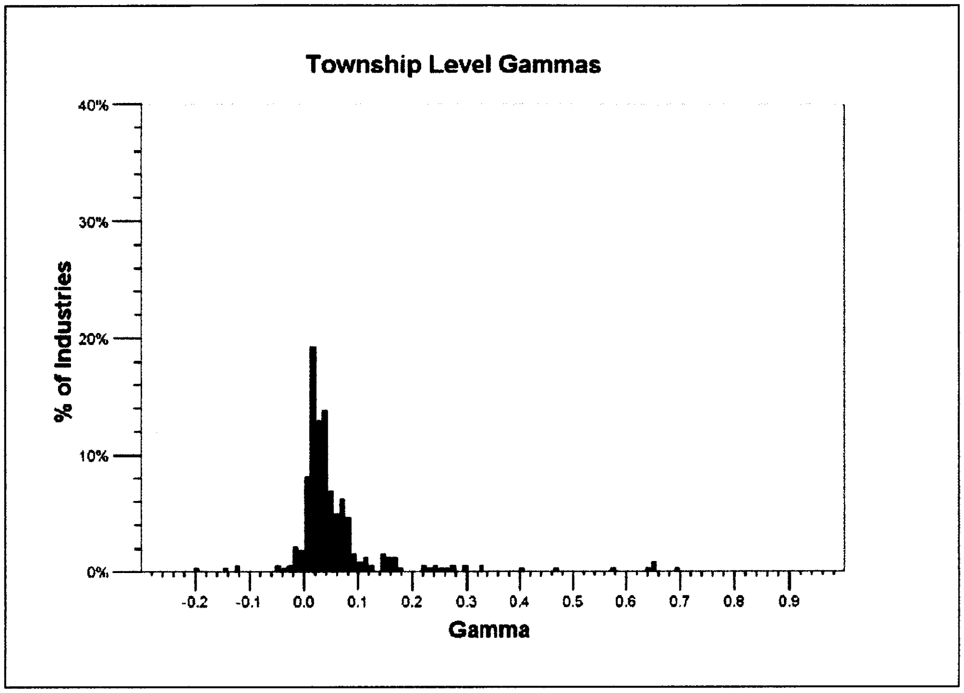


Figure 3-4 Histogram of γ at Town Level

As Rosenthal and Strange find in their study that measurement of EG concentration index becomes larger when data spatial unit changes from zip code to county level, and from county to state level, this is also witnessed in our study. In Table 3-3, both the mean and median of the EG index measurements increase from group to tract level and from tract to town level. This is also vividly displayed by three charts (Fig 3-2, 3-3, 3-4), with the decrease of the spatial scales, we see a rising degree of right skewness. This casts doubt on the existence of modifiable area unit problem (MAUP) within our data. To validate our guess, we carry out a comparison studies later this section with the use of artificial zones, which are different in scale, shape and position.

In our study, we calculate the values of three frequently adopted indices: Herfindahl index, spatial Gini index and EG concentration index. In a sense we hope these different measurements can coincide with each other and double check our calculation. Although we observe an agreed tendency of deviation from zero, high correlations among the measurements of same index at different scales (except Herfindahl), and high correlations between Herfindahl index and spatial Gini index, we find that the correlation between the EG concentration index and Herfindahl index, the EG concentration index and the spatial Gini index is strikingly low and is even with a reversed direction, as shown in Table 3-3. This finding conflicts with our common sense that these three indexes should give coincided information, at least in terms of direction. Ellison and Glaeser also notice the differences between spatial Gini index and EG concentration index, and they argue that "*some clusters appear at random accounts for at least as large a part of measured raw concentration as actual agglomerations of plants do. It is, therefore, not surprising that our index gives a different picture of firm concentration than previous discussions of raw concentrations have.*" While this argument works as an effective explanation for these differences between the EG concentration index and the spatial Gini, the Herfindahl indexes, it is still difficult to believe that the power of randomness is so large that it inverses the actual concentration (measured by the EG concentration index, and determined by natural advantages and spillover) from observed concentration (measured by the spatial Gini index and the Herfindahl index, and determined by natural advantages, spillover and randomness). Compared with the simplicity and straightforwardness of the Herfindahl index and the spatial Gini index, the EG concentration index is based on multiple assumptions and complicated deductions. One obvious limitation of Ellison and Glaeser location choice model is that it cannot explain why firms do not concentrate in one zone, which has the largest natural advantage, to enjoy the maximum spillover as well as maximum natural advantage, since it does not have punishment components for over-concentration, and neither does it have size restrictions.

- **Which industries are concentrated?**

After examining the overall concentration of all industries, we need to find out which industries are concentrated. Specifically, we may want to know those industries with the highest measurement values and those industries that are regarded as most important ones in Boston.

Table 3-4 and 3-5 list the top 10 industries with the highest measurements in terms of the EG concentration index and the spatial Gini index at three geographical levels: group, tract and township. Within each table, those industries that do not appear in list at next smaller geographical scale are marked with bold font. Both EG concentration index and spatial Gini index display strong coherence at different geographical scales. For the EG concentration index, only two industries at tract level: 3914 and 3531 do not appear at group level; and only three at town level: 3669, 2836 and 2211 do not show up at tract level. For the spatial Gini index, the most concentrated industries stay almost constant with a little change: 3914 at tract level but not at group level, and 2329 at town level but not at tract level. Although there exist similarities, the EG concentration index and the spatial Gini index identify different most concentrated industries. Industries in Table 3-4 consist of two major groups: food related (SIC code: 20) and industrial machinery (SIC code: 35). However, in Table 3-5, we see a mixed industries of food related products, chemical products (SIC code: 28), and laboratory related products (SIC code: 38).

Table 3-6 lists the top 10 industries with the largest employment, which includes: Surgical and Medical Instruments and Apparatus (3814), Electronic Computers (3571), Semiconductors and Related Devices (3674), Pharmaceutical Preparations (2834), etc. While these industries have medium to high spatial Gini index values, their EG concentration index measures are all strikingly low (all are negative values), which represent no concentration exist among these industries.

Table 3-4 Top 10 Industries with the Largest EG Concentration Index

Group Level			Tract Level			Township Level		
SIC	**SIC description**	γ	**SIC**	**SIC Description**	Γ	**SIC4**	**SIC description**	γ
3561	Pumps and Pumping Equipment	0.2618	3914	Silverware, Plated Ware, Stainless Steel Ware	0.5678	3914	Silverware, Plated Ware, and Stainless Steel Ware	0.5770
2011	Meat Packing Plants	0.1532	3561	Pumps and Pumping Equipment	0.2633	3561	Pumps and Pumping Equipment	0.2626
3452	Bolts, Nuts, Screws, Rivets, and Washers	0.1208	2011	Meat Packing Plants	0.1635	3511	Steam, Gas, Hydraulic Turbines, and Turbine Generator Set Units	0.2455
3911	Jewelry, Precious Metal	0.0833	3452	Bolts, Nuts, Screws, Rivets, Washers	0.1214	2011	Meat Packing Plants	0.1636
2098	Macaroni, Spaghetti, Vermicelli, and Noodles	0.0812	3531	Construction Machinery and Equipment	0.0964	3452	Bolts, Nuts, Screws, Rivets, and Washers	0.1250
3555	Printing Trades Machinery and Equipment	0.0643	3511	Steam, Gas, and Hydraulic Turbines, and Turbine Generator Set Units	0.0960	3669	Communications Equipment, Not Elsewhere Classified	0.1157
3511	Steam, Gas, and Hydraulic Turbines, and Turbine Generator Set Units	0.0610	2098	Macaroni, Spaghetti, Vermicelli, and Noodles	0.0832	2836	Biological Products, Except Diagnostic Substances	0.1085
2064	Candy and Other Confectionery Products	0.0568	3911	Jewelry, Precious Metal	0.0789	3531	Construction Machinery and Equipment	0.0978
3567	Industrial Process Furnaces and Ovens	0.0560	3567	Industrial Process Furnaces and Ovens	0.0594	2098	Macaroni, Spaghetti, Vermicelli, and Noodles	0.0869
2092	Prepared Fresh or Frozen Fish and Seafoods	0.0545	2064	Candy and Other Confectionery Products	0.0587	2211	Broadwoven Fabric Mills, and Cotton	0.0855

Table 3-5 Top 10 Industries with the Largest Spatial Gini Index

Group Level			Tract Level			Township Level		
SIC	SIC Description	Spatial Gini	SIC	SIC Description	Spatial Gini	SIC	SIC Description	Spatial Gini
3511	Steam, Gas, and Hydraulic Turbines, and Turbine Generator Set Units	0.8552	3511	Steam, Gas, and Hydraulic Turbines, and Turbine Generator Set Units	0.8578	3511	Steam, Gas, Hydraulic Turbines, and Turbine Generator Set Units	0.8405
3669	Communications Equipment, Not Elsewhere Classified	0.8398	3669	Communications Equipment, Not Elsewhere Classified	0.8423	2211	Broadwoven Fabric Mills, Cotton	0.8156
2836	Biological Products, Except Diagnostic Substances	0.8364	2836	Biological Products, Except Diagnostic Substances	0.8384	3669	Communications Equipment, Not Elsewhere Classified	0.8140
2211	Broadwoven Fabric Mills, Cotton	0.8085	2211	Broadwoven Fabric Mills, Cotton	0.8110	3812	Search, Detection, Navigation, Guidance, Aeronautical, and Nautical Systems and Instruments	0.8083
3812	Search, Detection, Navigation, Guidance, Aeronautical, and Nautical Systems and Instruments	0.7974	3812	Search, Detection, Navigation, Guidance, Aeronautical, and Nautical Systems and Instruments	0.7988	2099	Food Preparations, Not Elsewhere Classified	0.7714
2841	Soap and Other Detergents, Except Specialty Cleaners	0.7588	2841	Soap and Other Detergents, Except Specialty Cleaners	0.7608	2841	Soap and Other Detergents, Except Specialty Cleaners	0.6823
2099	Food Preparations, Not Elsewhere Classified	0.7525	2099	Food Preparations, Not Elsewhere Classified	0.7545	3821	Laboratory Apparatus and Furniture	0.6738
3821	Laboratory Apparatus and Furniture	0.6509	3914	Silverware, Plated Ware, and Stainless Steel Ware	0.7359	2759	Commercial Printing, Not Elsewhere Classified	0.6544
2759	Commercial Printing, Not Elsewhere Classified	0.6351	3821	Laboratory Apparatus and Furniture	0.6532	2836	Biological Products, Except Diagnostic Substances	0.6534
2084	Wines, Brandy, and Brandy Spirits	0.5425	2759	Commercial Printing, Not Elsewhere Classified	0.6371	2329	Men's and Boys' Clothing, Not Elsewhere Classified	0.6495

Table 3-6 Measurements for the Top 10 Industries with the Largest Employment

SIC	Description	Number of Establishments	Total Employment	Herfindahl	Spatial Gini Index			EG Concentration Index		
					Group	Tract	Town	Group	Tract	Town
3841	Surgical and Medical Instruments and Apparatus	273	48,075	0.1322	0.1200	0.1135	0.1151	-0.0124	-0.0195	-0.0154
3571	Electronic Computers	46	47,794	0.1801	0.1609	0.1620	0.1712	-0.0210	-0.0191	-0.0132
3674	Semiconductors and Related Devices	104	34,639	0.2508	0.2282	0.2070	0.2112	-0.0264	-0.0542	-0.0450
2711	Newspapers: Publishing, or Publishing and Printing	213	26,916	0.2790	0.2524	0.2545	0.2284	-0.0324	-0.0286	-0.0184
2834	Pharmaceutical Preparations	54	25,245	0.3171	0.2987	0.3026	0.3264	-0.0215	-0.0144	-0.0026
2741	Miscellaneous Publishing	273	24,649	0.3308	0.3122	0.3135	0.3192	-0.0219	-0.0187	-0.0060
2752	Commercial Printing, Lithographic	792	22,469	0.2007	0.1955	0.1970	0.1905	-0.0034	-0.0009	0.0055
2731	Books: Publishing, or Publishing and Printing	173	14,243	0.4983	0.4845	0.4867	0.5054	-0.0153	-0.0083	0.0168
2759	Commercial Printing, Not Elsewhere Classified	268	12,397	0.6514	0.6351	0.6371	0.6544	-0.0238	-0.0131	0.0341
3821	Laboratory Apparatus and Furniture	15	12,312	0.6683	0.6509	0.6532	0.6738	-0.0278	-0.0153	0.0355

b) Spatial Indices

In above paragraphs in this section, we have discussed the findings from non-spatial indices and related issues. We can see that these indices share a common feature towards industry concentration: viewing it as a static phenomenon that is measured at a single scale. Whatever detailed data are adopted, in these studies they are eventually aggregated to a certain spatial scale in order to develop an index representing the degree of industry concentration in the whole region. Nevertheless, in studies, researchers have noticed the existence of agglomeration at different scales. For example, when Rosenthal and Strange find obvious changes of EG concentration index when moving spatial unit from zip code to county level and from count to state level, they explain these changes by different county level and state level agglomeration phenomenon (although we argue in previous paragraphs that this is at least partly caused by MAUP). Marcon and Puech (2003) point out that while measuring industry concentration at single scale not only causes the result to be spatial unit dependent, but miss another important feature of agglomeration: the spatial structure of industry concentration at different scales at the same time. How does industry clustering vary with the scales in the study region? This is the question we will discuss in this section.

- **The overall concentration of manufacturing industry**

To find out at which scales the distribution of manufacturing firm establishments deviate from a random distribution, we compute the D function by differing the k value of manufacturing industries and that of the random sample mean, which comes from the 99 randomly generated distributions of the same size of points: 7828. The plot of the D function is shown in Fig. 3-5. In the chart, positive values of the D function mean manufacturing establishments are more concentrated than random samples and negative ones represent the contrary. The two dotted lines represent the maximum and minimum D values we can find using the 99 generated random samples and this gives us 99% confidence limits to reference our D values statistically.

To avoid too much edge effect, in the Fig. 3-5 we only show the computation of the D function within 10km range, which is almost half of the radius of the study region (24km). From the figure, we can see that the concentration of manufacturing firms appears at most of scales except 2.5 – 3.5km and 8 – 8.7km range. The peaks of concentration of manufacturing firm establishments look like a ripple starting from 1km. The significant concentration happens at 1-2km and 4.5-7km at 1% level. That is to say, the strong concentrations manufacturing firms at these scales are hardly caused only by random chances.

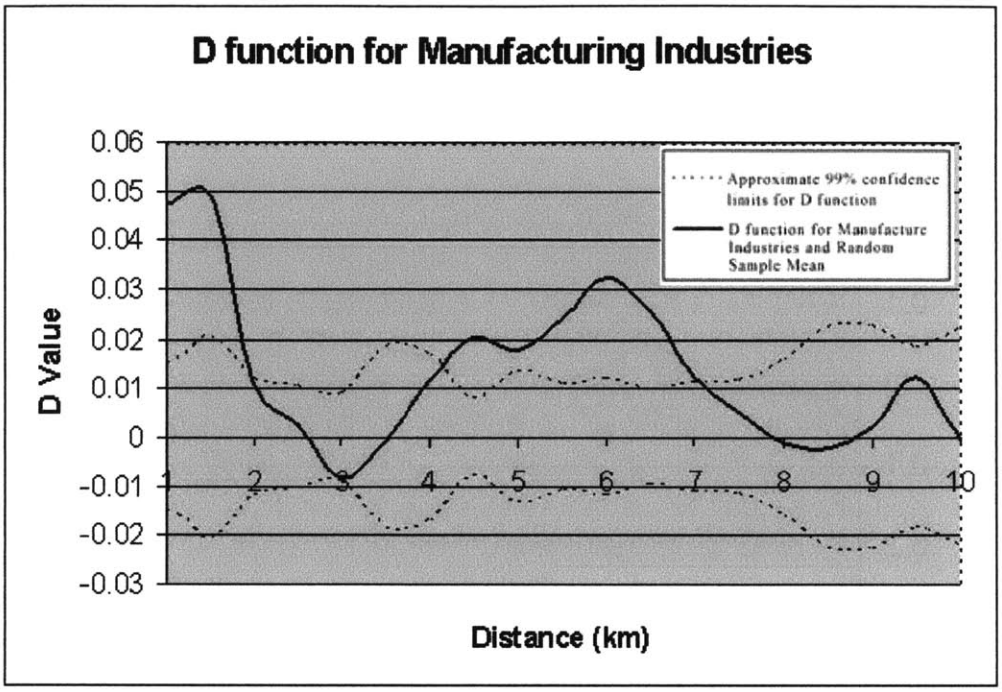


Figure 3-5 Plot of D Function for Manufacturing Industries

In addition to the concentration of manufacturing firms relative to random distributions, we may also want to find how manufacturing industries compare to other industries. In Fig 3-6, we plot the D function showing the relative concentration of manufacturing industries with respect to all industries in the study region. In theory all positive values mean that manufacturing industries are more concentrated than all industries in general at corresponding scales, and negative ones mean the contrary. While we can find from the figure that manufacturing industries are more concentrated than general industries at medium to large scales (2-4km, 5-10km) and more dispersed at small scales (<2km), these relationship are not that statistically significant. Almost the whole curve fall within the 99% confidence limits.

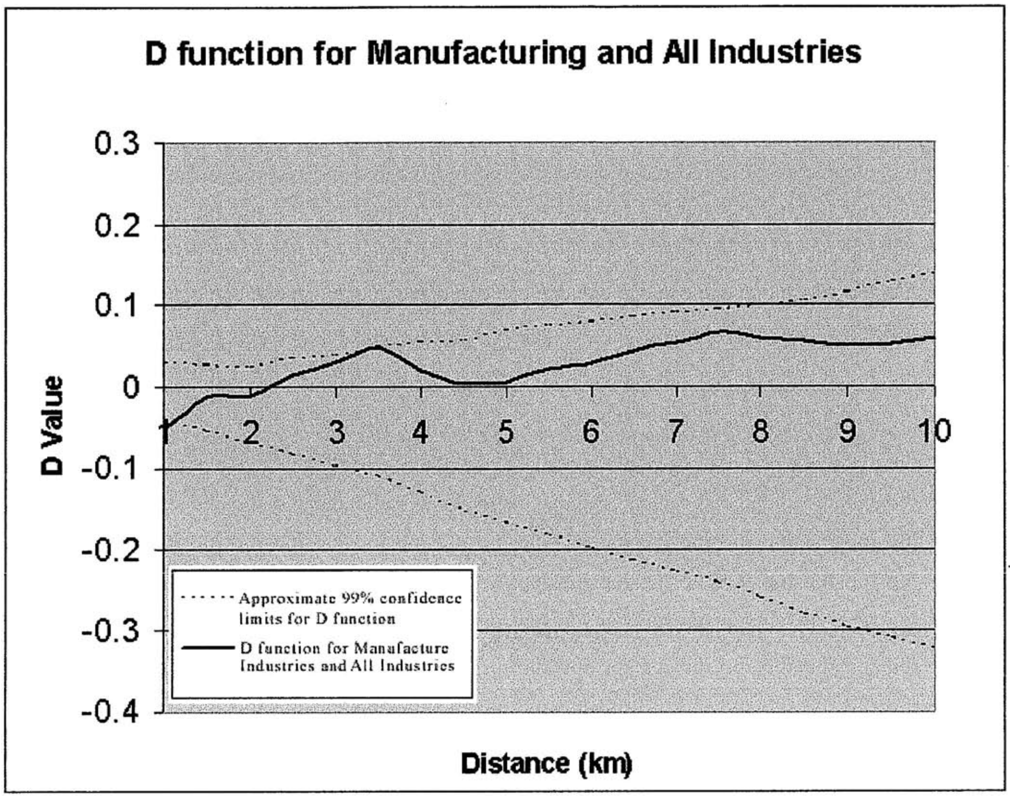


Figure 3-6 Plot of D Function for Manufacturing and All Industries

- **The sub-sectors in manufacturing industry**

After exploring the changes of the whole industry's concentration with changes of geographic scales, we go further to examine the characteristics of sub-sectors in manufacturing industry. Fig 3-7 shows the plot of the D function of "food and kindred products" industry, compared with the mean of 99 random samples. The solid and dotted lines represent the observed D function and the maximum D function from the samples respectively. The figure shows that "food and kindred products" industry has a little concentration within the range of 0 -3km and beyond that, this industry is actually dispersing. The industry of "Industrial and commercial machinery and computer equipment" has the contrary distribution to that of "food and kindred products" industry as shown in Fig. 3-8. From the figure, we can see that the industry of "Industrial and commercial machinery and computer equipment" display significant concentration at most of the geographic scales except those smaller than 2.2 km Industry of "Mesr/Anlyz/Cntrl Instrmnts; Photo/Med/Opt Gds; Watchs/Clocks" has a similar distribution of concentration with the industry of "Industrial and commercial machinery and computer equipment" (Fig. 3-9).

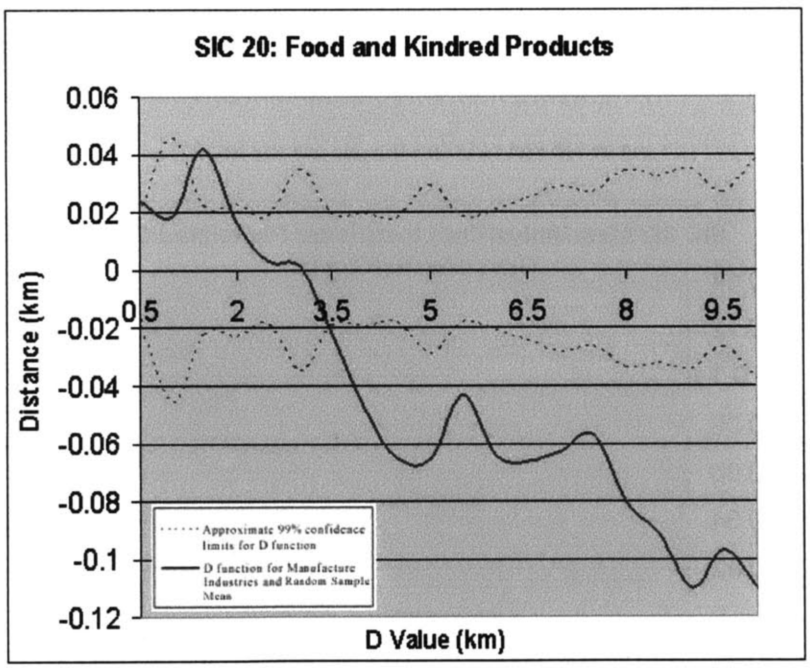


Figure 3-7 Plot of D Function for Industry SIC 20

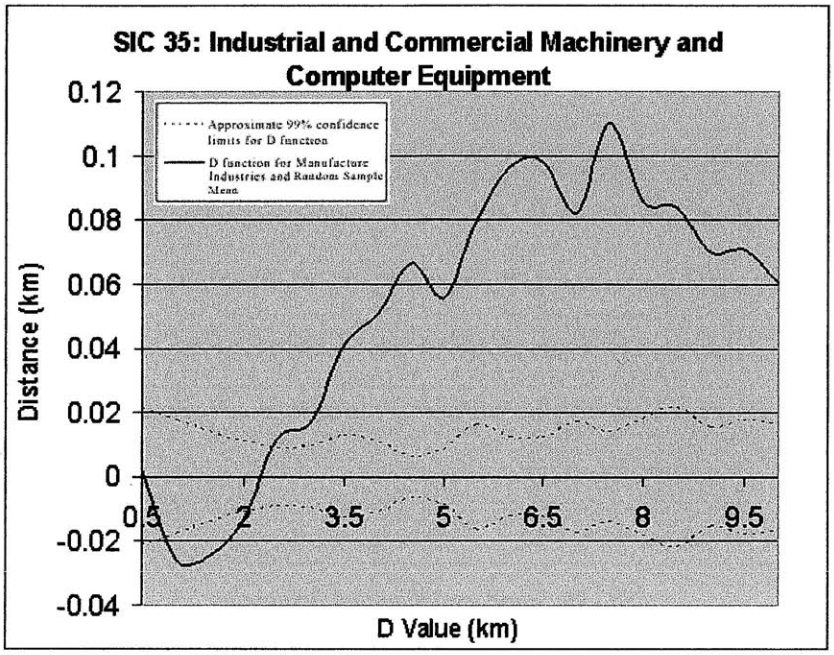


Figure 3-8 Plot of D Function for Industry SIC 35

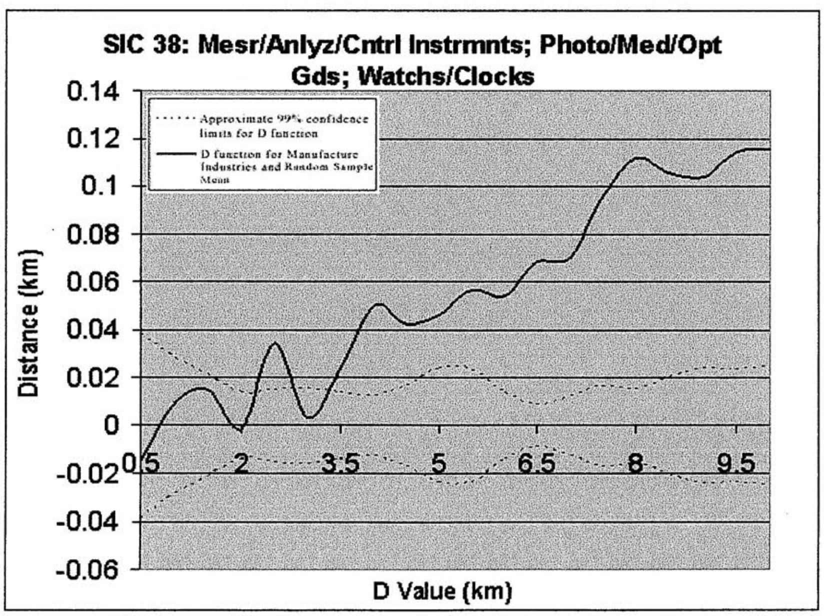


Figure 3-9 Plot of D Function for Industry SIC 38

3.4. Conclusion

This chapter reviews existing literature about industry clustering and discusses two categories of concentration indices: non-spatial index and spatial index. With the use of firm establishment data in Boston metropolitan area, we compute three indices: spatial Gini index, Ellison-Glaeser concentration index and the D function to compare their effectiveness in measuring concentration and integrate findings from these different perspectives into a relative comprehensive image of industry clustering. We also discuss computation related issues, especially MAUP.

Through measuring industry clustering in Boston, we witness ubiquitous existence of industry clustering in manufacturing industry, as previous studies have done, from the results of three indices adopted. We also identify an obvious inconsistency in the results between the spatial Gini index and the EG index in results is identified, as represented by low correlation coefficients and not agreed most concentrated industries. We doubt about the explanation provided by Ellison and Glaeser (2001) that the cause is that the EG concentration index measures concentration with randomness controlled for but the spatial Gini index does not, and believe the disparity between the EG concentration index and the spatial Gini index discloses the natural limitation of EG concentration index's model and computation.

The results of the D function show that spatial concentration of an industry varies significantly across different geographical scales within the study region. For manufacturing industry, we observe a significant concentration is observed at scales of 2-

4km, 5-10km and <2km and none significant dispersion is identified, when comparing the distribution of manufacturing firms with spatial random distributions at 99% significance level. However, there is no signal of significant concentration or dispersion when comparing manufacturing industry with all other industries. A further look at the concentration of sub-sectors in manufacturing industry shows that concentration of sub-sector industries also varies with the change of geographic scales, and technology intensive industries tend to be more concentrated than the others.

While this chapter tells us that industry clustering has local disfference and is spatially continuous in nature, some naturally followed questions to ask would be: at which locations do firms concentrate? What are the spatial scales of industry clusters? Why do firms choose to locate within or outside industry clusters? ... Obviously, these questions are beyond the capacities of the methods discussed in the chapter. We are going to explore these questions with new methods in the following chapter.

CHAPTER 4: DETECTING FIRM CLUSTERS OF URBANIZATION

4.1 Introduction

In Chapter 3, we measured and analyzed industry clustering with a variety of spatial and non-spatial indices. While these measurements give us an overall image about the industry clustering in the study region, there exist several limits in what they can tell about firm clustering. First, as the calculation is aggregated across the whole study region, the degree of industry clustering will be diluted when firm dispersion exists within the study area. The measurement results ultimately depend upon the selection of the study region, and we often observe the existence of Modifiable Area Unit Problem (MAUP). Second, the calculation of such industry clustering indices as the spatial Gini index and the EG concentration index requires data in certain aggregated formats, which are usually administrative district boundaries. As these subjective divisions of economic activities have no relevance to the underlying economic process which generates the industry clustering, there exists inherent error associated with these measurement results from the manipulated data, especially when the study area is relatively small. Third, industry clustering at pre-set regional levels provides no details about the locations and extents of firm clusters and leaves the following critical questions unanswered: Where do firms cluster? What are the spatial scales of firm clusters? Which firms cluster and how does their clustering behavior relate to their industry and firm characteristics?

In this chapter, we are going to discuss the method of detecting urbanization clusters, which are defined as the places with significant spatial concentration of industrial firms. As these clusters often appear in urban centers, where industrial firms coexist with other types of firms and urbanization economics are expected to play an important explanatory role for their existence, we call them clusters of urbanization or urbanization clusters to differentiate from other types of firm clusters. The major objects in this chapter are to find the locations, extents and statistical inferences of urbanization clusters. We will use Boston metropolitan area as a case study to implement and test the methods of detecting urbanization clusters. Specifically, using industrial firm establishment data, we will try to identify the urbanization clusters' locations, spatial extents and their statistical significance from spatial randomness in the study region. We will also investigate the clustering patterns of different firms, specifically knowledge intensive firms vs. labor intensive firms, and large firms vs. small firms.

4.2 Literature Review

4.2.1 Detection of firm clusters

Firm clustering within a region has been explained by two categories of economic externalities: urbanization and localization. Urbanization is an external economy of scale that can reduce the cost of doing business when firms, not specific to particular industry, stay together. Urbanization reduces the cost of inputs because of the low transportation cost of delivering inputs among firms, attraction of well-educated workers, and sharing of a wide range of business services (accounting, legal, consulting, etc) (Mittelstaedt &

Ward, 2006; Myrdal, 1957; Arthur, 1990). Localization was first discussed by Alfred Marshall (1890) and refers to another external economy of scale which occurs when firms in an industry locate close to one another. Three major advantages of localization are widely recognized and discussed: knowledge spillover, labor pooling and input sharing (Berry et al., 1997; Ellison & Glaeser, 1997; Rosenthal & Strange, 2001, etc.). In this study, the spatial concentration of industrial firms without controlling for unevenly distributed population is called *clusters of urbanization* or *urbanization clusters,* or just *clusters*, as they often appear in urbanized centers, where urbanization economies plays an important role. For those significant clusters with unevenly distribution of all firms controlled for, they are called *clusters of localization* or *localization clusters*, where localization economies are expected to plays an important role. As we can see urbanization cluster and localization cluster are not exclusive of each other. On the contrary, urbanization clusters can contain localization clusters, and vice versa. The core difference between them is that localization clusters are with exclusive significant concentration of industrial firms, compared with other firms, but urbanization clusters are just with significant concentration of industrial firms, without considering other firms. Although by default we refer to firm clustering as the spatial clustering of firm establishments, in many cases the understanding of employment concentration is our interest. Thus thereafter, we call the spatial clusters of firm establishments *clusters of establishment*, or just *cluster*; we call the spatial clusters of firm employments *clusters of employment*.

Early research has tried to detect firm clustering with purely statistical methods. For example, Rogers' study of retail establishment clustering with stochastic models (1965). However, when there is no prior knowledge about the location and scale of firm clustering within the study area (this is frequently the case), it is often required to scan the whole study area and test all possible clusters to find those true clusters. Purely statistical methods are proved to be unable to provide either accurate or unbiased answers (Croasde and White, 1987) due to a few technical difficulties, which includes selection bias when selecting specific time periods or sub-regions of the data, heavy reliance on the prior knowledge about clusters in formulating hypotheses and problems associated with determining the significance of results, etc (Openshaw et al., 1987).

A traditional approach to detect point clusters is to sample the study area using a system of quadrants. This is usually done by exhaustively partitioning the whole study area (Choynowski, 1959; Greig-Smith,1964; Kershaw, 1964). The resulting observed distribution of cell counts is then compared to a model derived from some theoretical spatial processes, such as the Poisson distribution. For example, in studying the distribution of brain-tumors in the Rzeszow Province in Poland, Choynowski (1959) treats the 17 counties within the study region as quadrants and counts the number of diseases within them. Those quadrants with diseases over a significance level α are regarded as the disease clusters. The significance level α is determined by the Monte Carlo method.

The Quadrant-based method provides a convenient solution for analyzing point patterns and is widely adopted in studies. However, it also suffers from several defects. First, the result is not scale-free and dependent upon the quadrant size used in the study. When the quadrant size changes, the result will also change (Peilou,1957). This causes selection bias. Second, as quadrants are usually fixed in the study, this method will fail to detect the exact locations and extents of clusters unless the clusters are much larger than the quadrant size. Third, a basic assumption in Poisson models, which are extensively used with this method, is that the counts for various quadrants are independent of each other. However, this can be true only when the quadrant size is significantly larger than the cluster size. Otherwise, the counts will be auto-correlated.

Openshaw (1987[1]) regarded that many of the problems of previous efforts in analyzing point patterns was the result of selectivity of the hypothesis being tested. Limited by the nature of purely statistical methods and traditional quadrant-based methods, only a few hypotheses could actually be raised and tested. For example: whether there exists concentration in area A or not, whether there exists concentration at scale B or not. However, without prior knowledge about the scales and locations of the clusters, numerous hypotheses need be constructed and tested before we can identify the clusters at different scales. Thus, Openshaw suggested that the most general and least-biased solution for detection of point clusters would be to generate and test all possible geographical hypotheses and he called the method: Geographic Analysis Machine (GAM) (Openshaw et al., 1987; Openshaw and Craft, 1988; Openshaw, 1989a, b). In implementation, Openshaw laid multiple overlapping circles of variable size as quadrants on the study region, counted events within each quadrant and tested it independently with a Monte Carlo process. In nature, Openshaw's method is identical to previous quadrant-based efforts like Choynowski's except that in Openshaw's method, the quadrants overlap and vary in size. With Openshaw's method, in theory we can scan the study region to find clusters at any scale and any location, although in practice, as restrained by computation time and capacity, only a few chosen hypotheses are constructed and tested.

GAM produces large impacts on spatial analysis of point patterns and has been widely referenced by subsequent research. However, it is also criticized for a few defects. The most severe one is that while it avoids the problem of spatial selection bias by testing all possible sub-regions, it also suffers from multiple testing (Besag et al. 1991). Turnbull et al. (1990) made two changes over Openshaw et al.'s method to solve the multiple testing problem. First, instead of making circles with same radius, Turnbull et al. drew circles at the radius to make them contain exactly the same population, which is called "population radius" by Kulldorff et al. (1994). Second, for each population radius, the biggest number of cases within all circles was tested by using the Monte Carlo method, instead of testing all counts within the circles. Turnbull et al. argued that the statistics constructed in this way would have identical distributions, but would not be dependent. However, to detect the clusters at different scales, Turnbull et al. suggested repeating their procedure with different values of population. Obviously, this reintroduces the problem of multiple testing.

Compared with GAM, Turnbull et al.'s method largely increases the computation complications, yet cannot be used to solve the problem. As our study in this section is to get a general image of the locations and scales of industry clusters, multiple testing will not cause much trouble for our study. In the following, we introduce the methodology and computation-related issues of GAM. By carrying out GAM in our study region, we analyze the locations and scales of industry clusters, and discuss how industry and firm characteristics affect industry clusters.

4.2.2 Firm clustering and firm size

There is evidence in the literature that firms gain productivity advantages from agglomeration (Henderson, 1986; Ciccone and Hall, 1996). If so, then plants within the clusters may grow faster as well and we might have a better chance to observe the concentration of large size firms compared to smaller size firms. Duranton and Puga (2001) provided another theory explaining the relationship between the size of the plant and their location choices. In their model, firms switch to mass production and relocate to specialized cities with lower production cost while optimizing their production. Thus larger firms show larger propensity to collocate in specialized areas. This is verified by the studies of Florence (1948) and Kim (1995). They identified positive relationships between the average plant size in an industry and its geographic concentration. By taking advantage of Ellision and Glaeser's geographic concentration index, Homes and Stevens (2002), Lafourcade and Mion (2003) further find that within specific industries, the large plants cluster more than the small ones. As it is commonly believed instead that smaller firms should be more clustered since they need knowledge-spillover (which is regarded as one of the major factors causing agglomeration) more than larger firms, Alsleben (2005) argued that the reason for observing large firms being more concentrated is that small firms are more easily affected by increasing labor cost and technology breaches thus they choose to stay away from industry clusters.

While these analysts concentrate on the clustering behaviors of firms of various sizes at a regional level, an equally important question is whether we will observe similar results within regions. Will large firms tend to locate within industry clusters more than small firms? After identifying significant industry clusters, we examine this question.

4.2.3 Firm clustering and production knowledge

It is widely accepted by researchers that firms benefit from the knowledge spillover which gives rise to productivity advantages, as Mittelstaedt and Ward (2006) found that firms gain production efficiency improvement when observing, copying (or stealing) the best workers and practices of others in their industry. Romer (1986) found the benefit that firms can obtain from knowledge spillover is sensitive to distance. This is especially obvious when the knowledge is tacit in nature. Empirically, Jaffe, Trajtenberg and Henderson (1993) tracked the knowledge flows in the form of patent citations, and found that patent citation is significantly more likely to occur within the same city. These studies suggest the significance of knowledge spillover as well as its spatial decay nature, which is regarded as one of the important agglomeration economies. Many other studies have been carried out to understand the underlying mechanism for knowledge spillover. Aydogan and Lyon (2004) argued that technological complementarities increase the benefit of cooperation in R&D. Varga (1998) identified that channels of the university knowledge transfer are: R&D cooperation between academia and industry; university seminars; scholarly journal publications; faculty consulting; industrial associates programs; industrial parks; spin-offs (faculty and students); technology licensing; the local labor market of scientists and engineers; and local professional associations of scientists.

While the benefits from knowledge spillover within industry clusters have been identified and understood from both empirical and theoretical studies, it is natural to presume that knowledge-intensive firms, which appreciate the value of knowledge more than others, should gain more benefit from spillovers, thus be more likely to cluster. However, several analysts (Glaeser et al, 1992; Henderson, 1997; Anselin et al., 1997; Ellison and Glaeser, 1997; Rosenthal and Strange, 2001) show that high-technology industries, in which spillover should be prevalent, show little clustering. As most of these studies are carried out at a regional level, again, we re-enter this topic to see whether knowledge dependency in production affects firms' locations with respect to industry clusters.

4.3 Detection and inference of urbanization clusters in Boston

4.3.1 Data

a) Firm establishment data

This empirical study will detect and infer manufacturing firm clusters in the Boston Metropolitan Area (BMA). The manufacturing firms are those with SIC codes between 20 and 39. BMA is defined as all townships within the I495 Highway. The total area of the BMA is about 5500 Km^2. While detecting clusters, the firms are differentiated by their spatial locations; if a firm has several branches within the BMA, each branch will be recognized as an independent observation. We use 2009 firm data at the establishment level from InfoUSA database in the study. Other than latitude and longitude values of firm establishments, InfoUSA also provides additional information including location employment, sales and SIC code, etc. In total, there are 7828 manufacturing firm establishments within the BMA.

b) Industry employee education attainment data

The education attainment data for industries in Table 4-1 comes from the Current Population Survey (CPS) 2006, provided by the Bureau of Labor Statistics (BLS). The CPS data reports the industry of occupation for individual workers by census industry codes. We converted the census industry codes 4-digit SIC codes according to a correspondence table provided at the census website.

(http://www.census.gov/hhes/www/ioindex/ioindex02/download02.html).

Table 4-1 Occupations with Top 20 Worker Education Attainment

Occupation	Percentage with Bachelor or higher degree
Scientific research and development services	0.74432
Pharmaceutical and medicine manufacturing	0.62532
Publishing, except newspapers and software	0.59259
Computer and peripheral equipment manufacturing	0.58491
Aerospace product and parts manufacturing	0.52510
Communications, audio, and video equipment manufacturing	0.49432
Sound recording industries	0.48649
Electronic component and product manufacturing, n.e.c.	0.48601
Navigational, measuring, electromedical, and control instrument	0.45339
Commercial and service industry machinery manufacturing	0.40594
Industrial and miscellaneous chemicals	0.39806
Medical equipment and supplies manufacturing	0.37654
Newspaper publishers	0.37317
Soap, cleaning compound, and cosmetics manufacturing	0.37162
Aircraft and parts manufacturing	0.35398
Paint, coating, and adhesive manufacturing	0.34783
Beverage manufacturing	0.34409
Agricultural chemical manufacturing	0.33333
Footwear manufacturing	0.30556
Pottery, ceramics, and related product manufacturing	0.30435

4.3.2. Methodology

a) <u>*Detect firm clusters of urbanization*</u>

The computation of GAM consists of three major steps: generating hypotheses, testing hypotheses and displaying significant clusters. The null hypothesis that GAM is set to test is whether there exists no point clustering of any size at any location. Without prior knowledge about the point distribution, this generic hypothesis needs to be constructed and tested at various scales of interest throughout the study region. To do this, we lay a regular lattice of points over the study region with spacing distance of z. Then we draw a circle of radius r by using each of the lattice points as a circle center. To ensure a similar degree of overlapping among circles, we constrain z to meet: $z=a*r$, where a is called overlapping parameter. For each of the generated numerous circular "zones", we count the firm establishments that fall in the zones and we use the count as a test statistic to be compared with some criteria value in order to determine whether firm clustering exists within the study circle. This hypothesis generation and testing process is repeated for all locations and scales to identify all potential industry clusters.

The criteria value analysts use for assessing the significance of the test statistic generated from each circular zone is achieved through different approaches in the studies. Some researchers use Poisson probabilities by assuming that the distribution of firm establishments under study is generated by a Poisson process (Craft et al., 1985). However this can be problematic since many processes are equifinal (Harvey, 1968).

That is to say more than one process can produce the same end pattern (Cliff and Ord, 1975). When inferring a generation process from the observed point pattern, we cannot get certainty from a good fit. Thus a more general approach to inferring the test statistic is achieved through the Monte Carlo procedure, which was developed by Hope (1968) and is widely applied in studies (Openshaw et al, 1987; Kulldorff and Nagarwalla, 1995; Goreaud, 2000; Marcon and Puech, 2003). The Monte Carlo procedure determines the significance level of the test statistic by producing a number of simulations of spatial distributions from the null hypothesis and these simulations have the same number of events as the study distribution. The top *n* percentage of the values of test statistic calculated from these simulations is chosen to be the lower bound of n% significance interval. For example, in our study, we use 100 repetitions of spatial random distributions of the 7828 points (the number of total manufacturing establishments in the study area) are produced, and for each of the simulations, we count the firm establishments with a circular zone of radius r. The biggest value of the count from all the samples represents the lower bound of 1% significance interval for scale of r. Once we determine the significant intervals by the Monte Carlo procedure, we identify those clusters at 1% significance level and mapped them with ArcGIS software.

b) <u>*Define large-size plants*</u>

To analyze the impacts of plant size on the industry clusters, we first partition the universe of firms into two groups (large and small). Then we are going to identify the significant industry clusters for large pants and small plants separately and discuss the impacts of plant size by comparing the two groups of identified clusters. The cut-off value for large firms is set to be 35 and larger, compared with 20 in Holmes and Stevens' study (2000). This gives us 2056 large firms, which account for 26% of the total manufacturing establishments.

c) *Define knowledge-intensive and labor-intensive firms*

The role of knowledge in the production of a plant is hard to capture since there are not any form of records about their existence and their flows among plants. Alsleben (2005) argued that knowledge should be incorporated in "key employees" (e.g. engineers, managing directors) and may flow between firms through labor poaching if firms locate close to each other. Rosenthal and Strange (2001) proxy the knowledge-intensity of employment with employment's education attainment, calculated as; the percentage of workers with Doctorates, Master's Degrees, and Bachelor's Degrees in the total employment of an industry. In this study, we will follow this strategy and regard a high percentage of employment with Bachelor and more advanced degrees as a symbol for strong knowledge dependency in the production process of the plant. All plants are divided into two groups: knowledge-based plants and labor-based plants using a cut-off value, which is set to be 50%. In our data, 2029 plants are identified as knowledge-based industries, and 5746 as belonging to labor-based industries. The summary information about this partition is shown as Table 4-2.

Table 4-2 Distribution of Industry's Worker Education Attainment

	Number	Mean	Std.	Definition	# of Establishments
All SIC Codes (4-digit)	368	0.17	0.076	All Available Data	7828 (53 SIC codes)
Knowledge-intensive industries	67	0.30	0.04	>=0.50	2029
Labor-intensive industries	301	0.15	0.04	<0.50	5746

4.3.3. Results

a) <u>*Significant clusters of urbanization within Boston Metro Area*</u>

Table 4-3 summarizes the number of clusters tested and the significant clusters found to be a statistical significant departure from the expected Poisson pattern for radii in the range of 0.5-10 km, with a 0.5 step increase. At the .01 level of significance, 68%-80% of firms are found to locate within industry clusters. This may have two explanations: first, in Boston, most of the manufacturing firms do cluster within clusters of certain sizes; second, it may also reflect a loose criteria has been applied in evaluating firm clusters, which produces the spurious omnipresence of firm clustering. Thus, we also calculate industry clustering at the .001 level of significance to re-examine the results. At the .001

level of significance, 38% - 67% of firms are found within the significant clusters, and at the radii between 1km and 10km, the percentage of firms found within significant clusters varies around 60%.

The small variance of percentage of firms found within significant clusters between 1km and 10km signifies the omnipresent firm clustering at different scales. Concerns may arise while adopting the lower-than-usual significance level of .001 as this criterion excludes some clusters from being identified. Because of the existence of multiple testing problem in GAM, applying an arbitrarily low significance level will reduce the presence of Type I errors in the findings and help focus attention on the most significant clusters (Openshaw, 1997).

Table 4-3 Distribution of Significant Clusters

Circle Radius (km)	Total Clusters Tested	P=0.01		P=0.001	
		Number of significant clusters	Percentage of firms within significant clusters	Number of significant clusters	Percentage of firms within significant clusters
0.5	139381	2132	0.49	1210	0.38
1	34,825	2,126	0.68	1473	0.58
1.5	15,471	1,352	0.69	901	0.60
2	8,710	893	0.71	598	0.63
2.5	5,579	684	0.75	409	0.64
3	3,865	494	0.77	258	0.61
3.5	2,844	405	0.78	208	0.62
4	2,173	276	0.77	157	0.56
4.5	1,718	230	0.78	125	0.54
5	1,390	208	0.79	111	0.54
5.5	1,143	195	0.80	102	0.59
6	964	163	0.75	92	0.58
6.5	827	133	0.76	79	0.56
7	708	132	0.75	73	0.55
7.5	617	120	0.78	62	0.53
8	548	116	0.80	62	0.65
8.5	484	104	0.80	53	0.54
9	431	91	0.80	51	0.57
9.5	381	85	0.81	49	0.66
10	349	78	0.80	40	0.67

Figure 4-1 gives the distribution of the existing 7828 manufacturing firm establishments and each dot represents one establishment. From Figure 4-1, we can learn that existing manufacturing establishments mainly concentrate on the central areas of Boston and scatter toward the north, south and west along major highways: North I-93, Highway 3, South I-95, Highway 90 and Route 128.

Figures 4-2 to 4-7 list the identified significant clusters of establishment and clusters of employment within 3km range. As shown in Figure 4-2, Boston downtown area is found with the most urbanization clusters. Outside the central area, obvious concentrations of urbanization clusters can also be found along the highway corridors of I-95, Route 128 and I90. Nearby Framingham, Norwood, Brockton and Weymouth, firm clusters gather around. The distribution of firm clusters or urbanization clearly follows a hub-and-spoke pattern with some important towns as the hubs and the connecting major highways as the spokes.

Comparison of clusters of establishment and clusters of employment shows that while the latter is more concentrated along the I90 and Route 128 corridors, the former is more scattered, but still follows a similar hub-and-spoke pattern as the latter does. This may be caused by the distribution of firms with large employment. As large firms are more concentrated in core clustering areas, they pull the distribution of employment into a more restrained pattern compared with that of firm establishments, which is mainly affected by small firms.

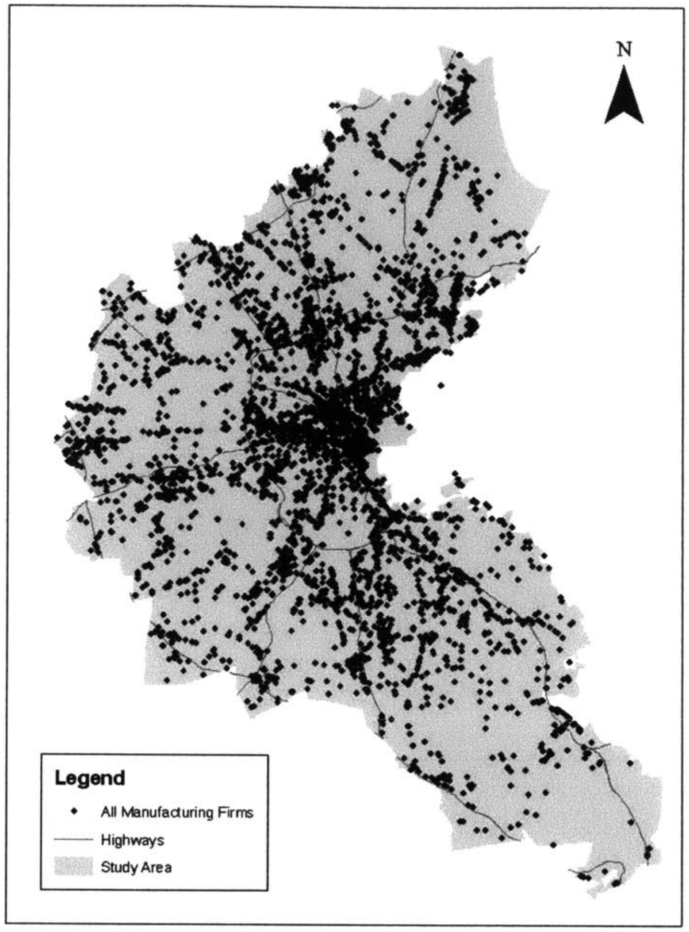


Figure 4-1 Location of Manufacturing Plants

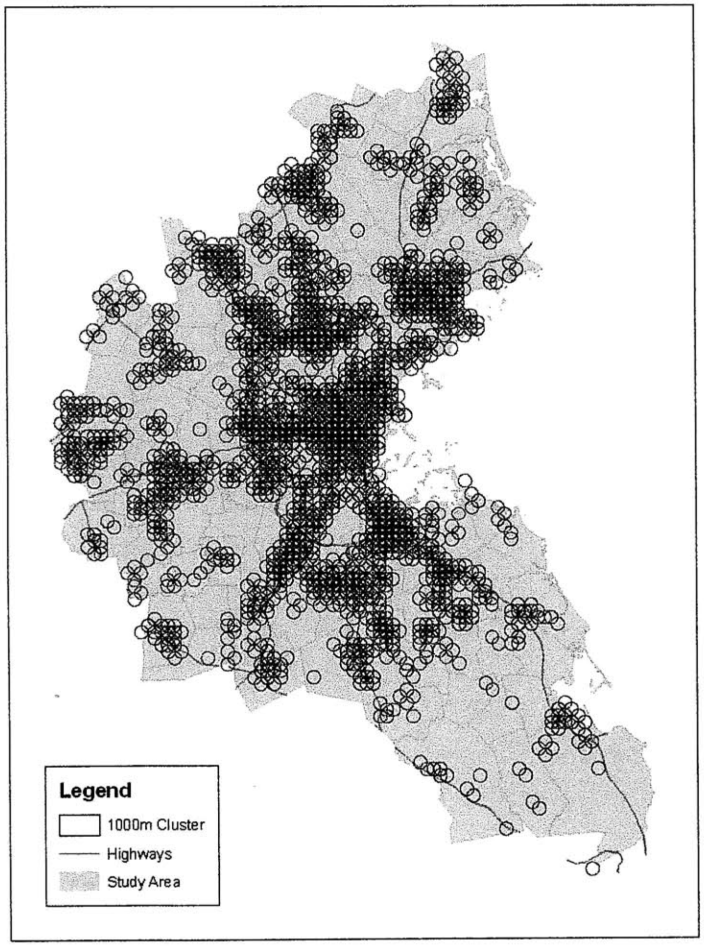


Figure 4-2 Clusters of Establishment at 1km Scale

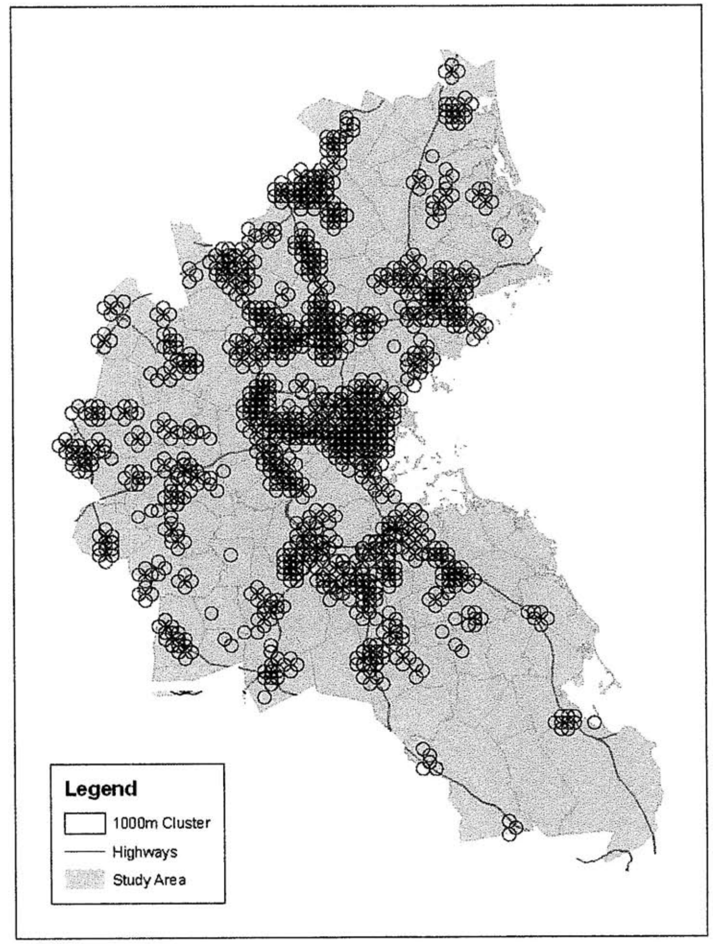


Figure 4-3 Clusters of Employment at 1km Scale

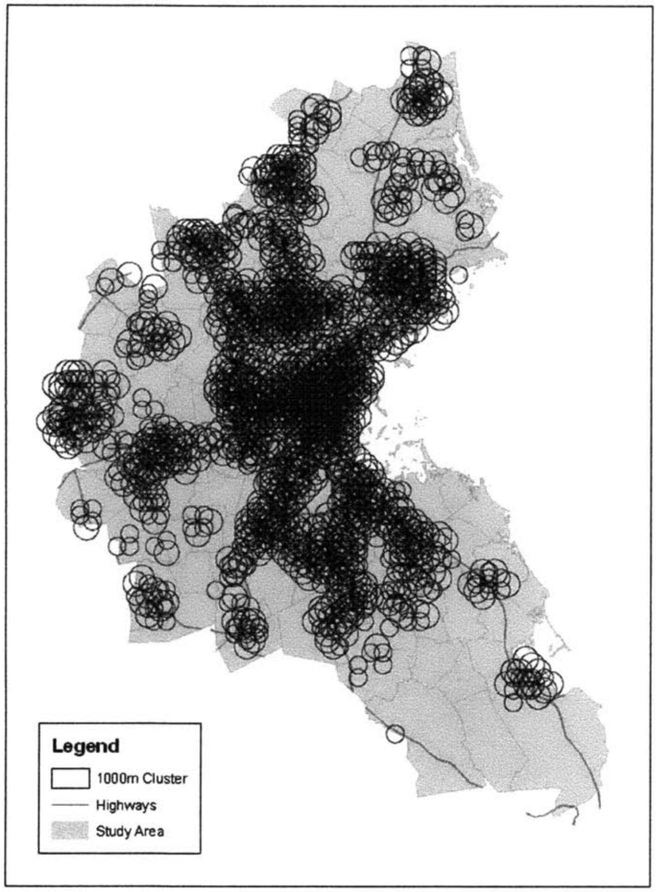


Figure 4-4 Clusters of Establishment at 1-2km Scale

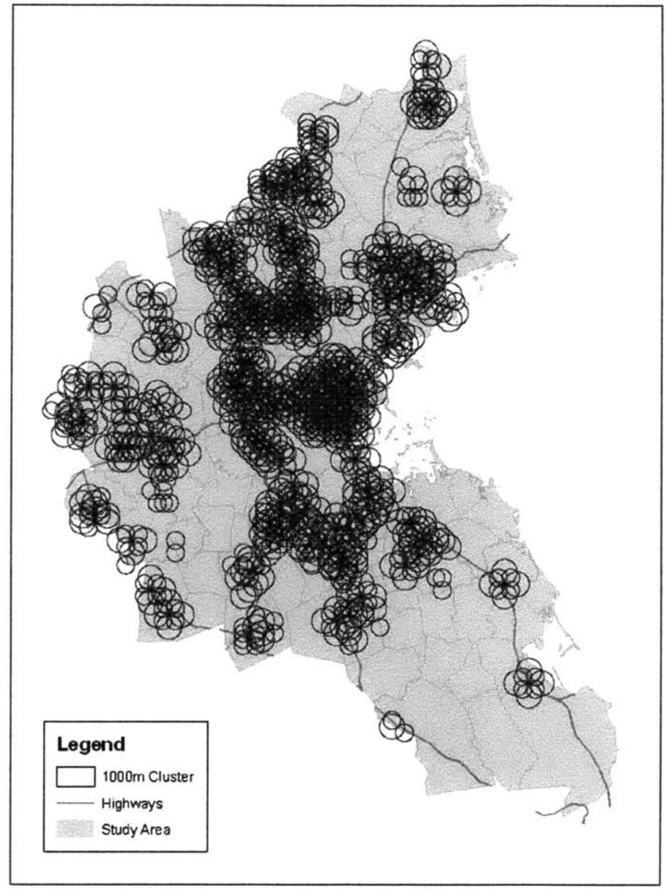


Figure 4-5 Clusters of Employment at 1-2km Scale

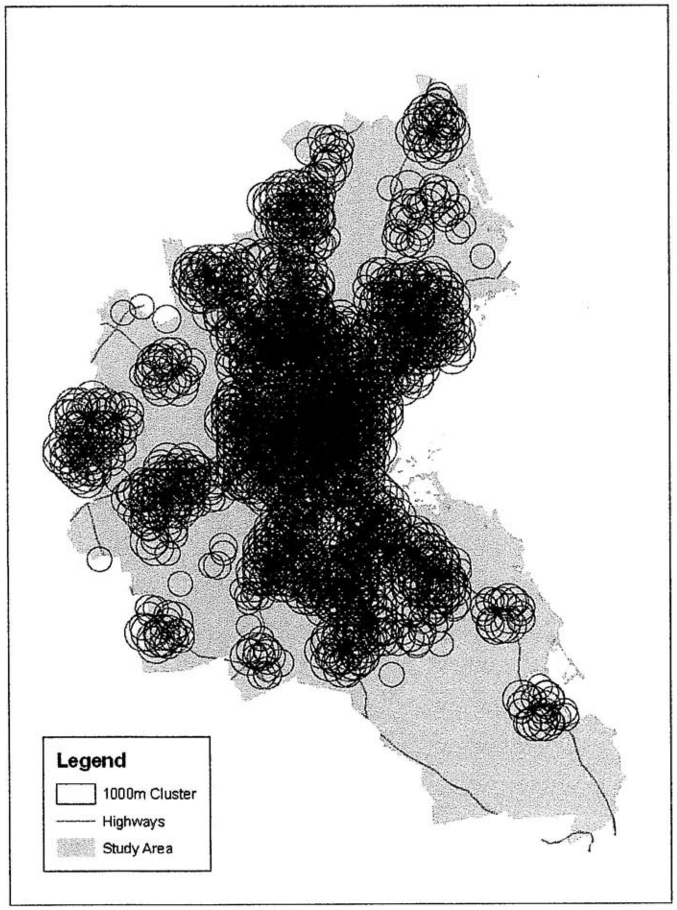


Figure 4-6 Clusters of Establishment at 2-3km Scale

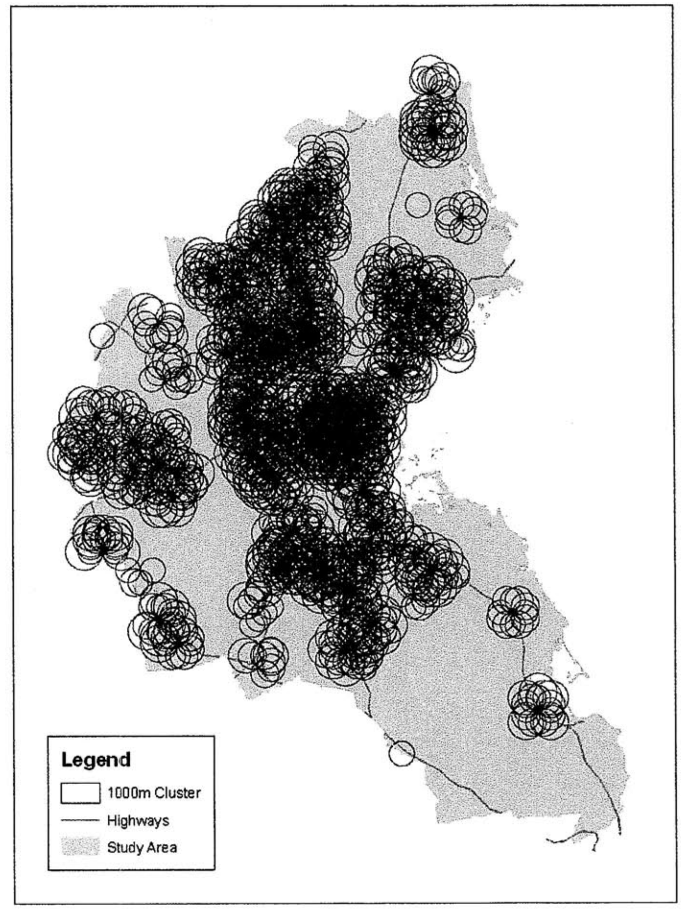


Figure 4-7 Clusters of Employment at 2-3km Scale

Another observation from above clustering maps is the overlap among clusters. The firm clusters at different scales overlap each other and constitute a complicated system. When a firm locates within a cluster, it usually falls within a series of overlapping clusters at various spatial scales. When we increase the spatial range from 1km to 3km, we observe a steady, instead of a sudden, changing pattern. That means while obviously space plays a role in firm's location choices, it plays a little differently at different scales.

b) Firm size and firm clustering

Figures 4-8 to 4-13 presents the significant clusters of large firms and small firms at the 0.001 level of significance within the range of 3km. As previous maps have shown, clusters of large firms and small firms are also found with obvious concentration in the downtown area. Outside of the central area, significant firm clusters distribute along the major highways of I-95, I-90, I-495 and Route 128. We can also observe that clusters of small firms are much more dispersed than those of large firms, which tightly distribute along highways and within downtown. This observation agrees with the literature findings that while large firms are with bigger bidding capacity for best locations, small firms have to look at wider areas for locations.

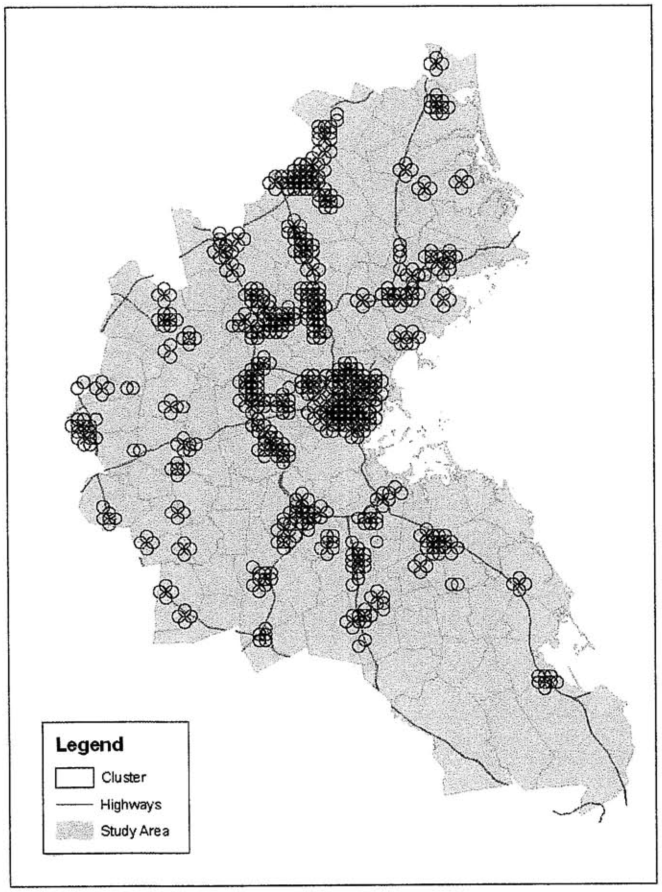


Figure 4-8 Clusters of Large Firms at 1km Scale

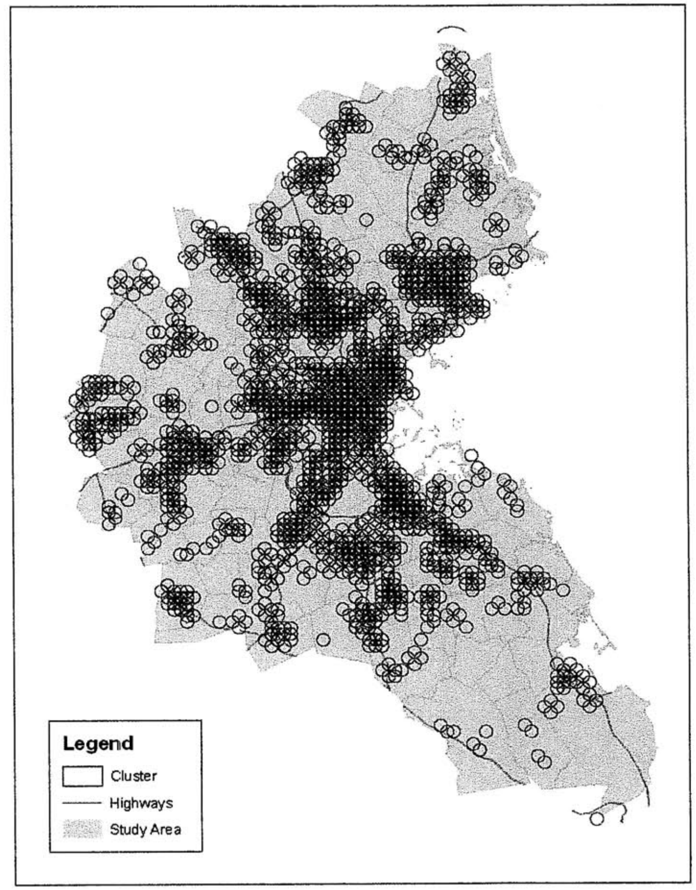


Figure 4-9 Clusters of Small Firms at 1km Scale

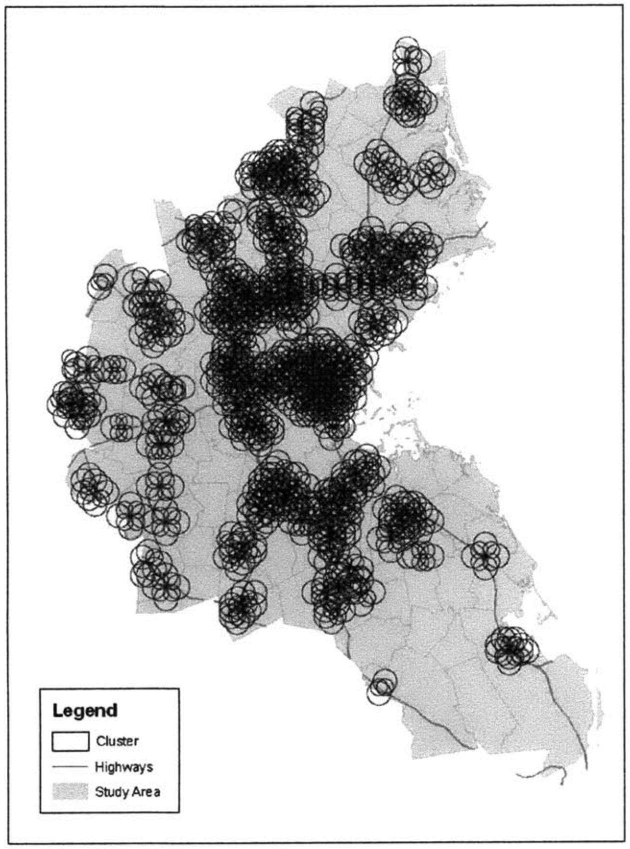


Figure 4-10 Clusters of Large Firms at 2-3km Scale

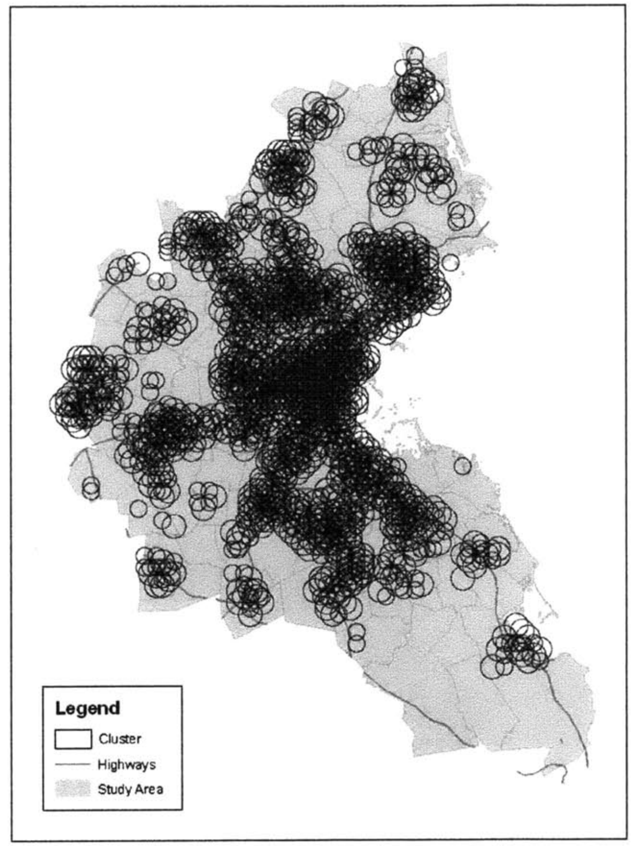


Figure 4-11 Clusters of Small Firms at 2-3km Scale

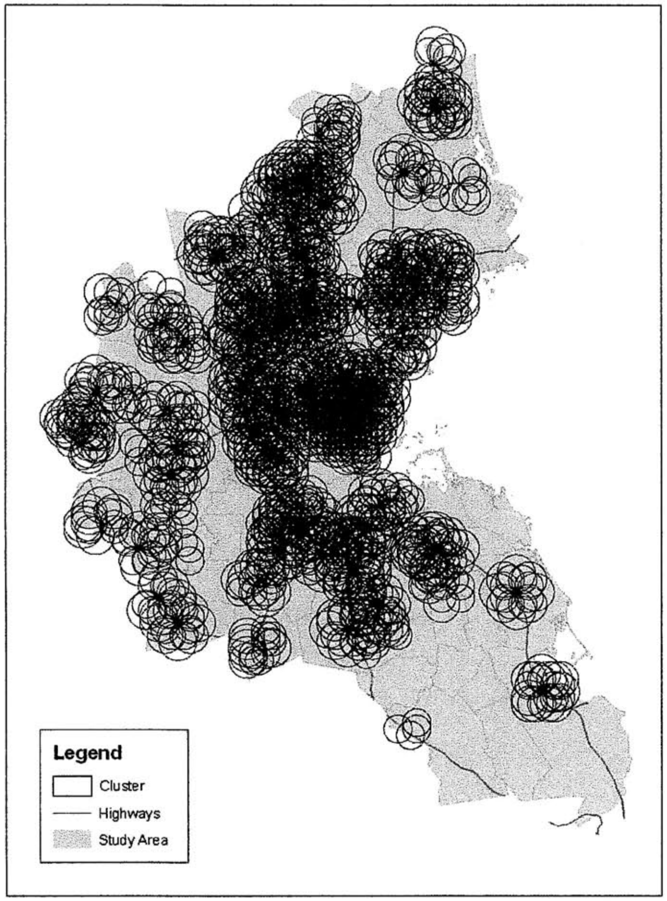


Figure 4-12 Clusters of Large Firms at 2-3km Scale

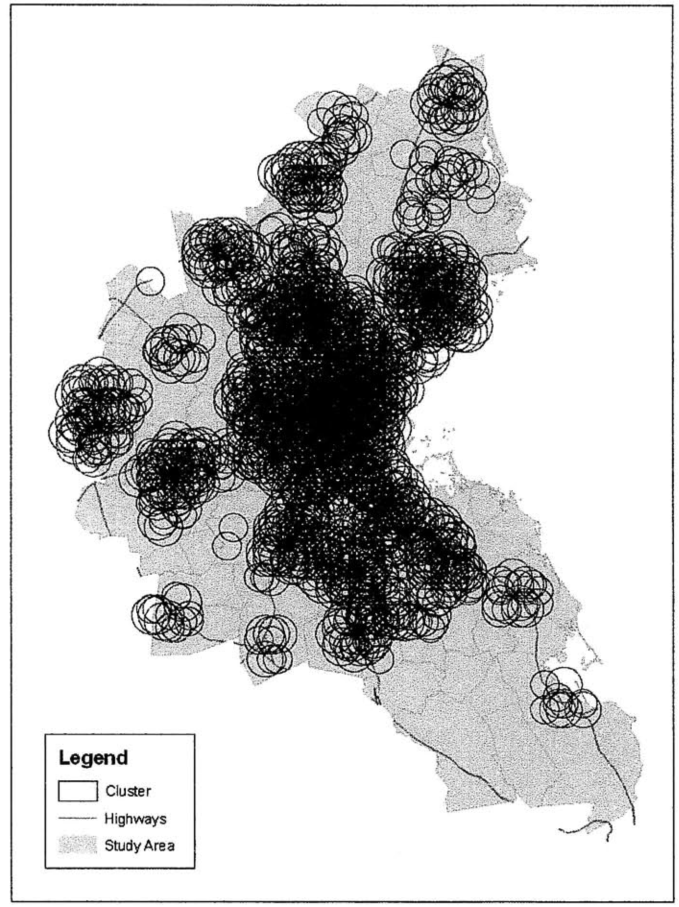


Figure 4-13 Clusters of Small Firms at 2-3km Scale

c) Firm's knowledge intensity and firm clustering

Figures 4-14 to 4-19 presents the identified significant firm clusters of labor-intensive firms and knowledge-intensive firms at the .001 level of significance within 3km range. At 1km scale, clusters of knowledge-intensive firms exhibit a more compact but less continuous pattern compared with the clusters of labor-intensive firms. At this level, locations like downtown, Norwood, Dedham, Woburn, Peabody and Framingham attract most of the knowledge-intensive firm clusters. At 2km range, clusters of knowledge-intensive firms start to be more continuously distributed and as dispersed as clusters of labor-intensive firms. From 2km to 3km, the biggest change is the clusters of knowledge-intensive firms become even more dispersed than those of labor-intensive firms.

Previous empirical studies by Glaeser et al (1992), Henderson (1997), Anselin et al. (1997), Ellison and Glaeser (1997), Rosenthal and Strange (2001) find that technical-intensive firms are less concentrated than labor-intensive ones. However, with the increase of range, at 3km scale, clusters of knowledge-intensive firm start to be more dispersed than clusters of labor-intensive firms, which mainly concentrate around downtown and Lawrence areas. What previous empirical studies have discovered are possibly some facts about firm clustering at large scale or regional level. As the pattern of firm clusters change with the spatial range, it is essential to study them at various levels to get a comprehensive view.

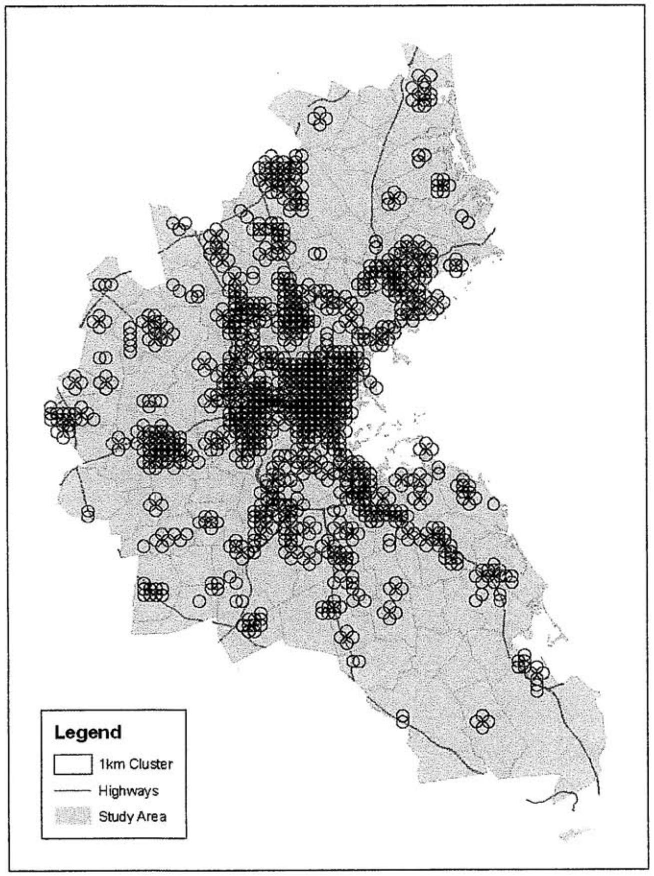


Figure 4-14 Clusters of Labor-intensive Firms at 1km Scale

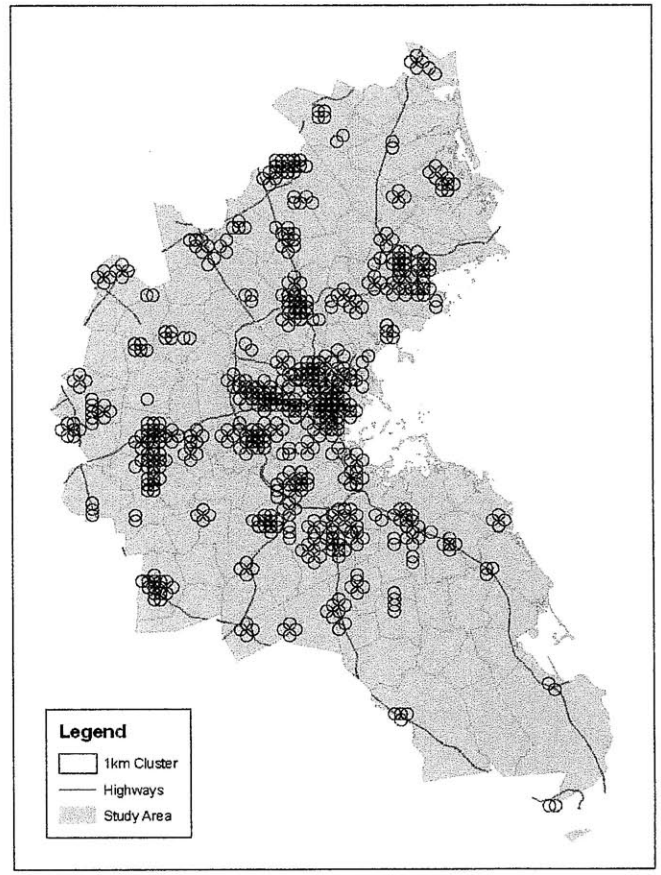


Figure 4-15 Clusters of Knowledge-intensive Firms at 1km Scale

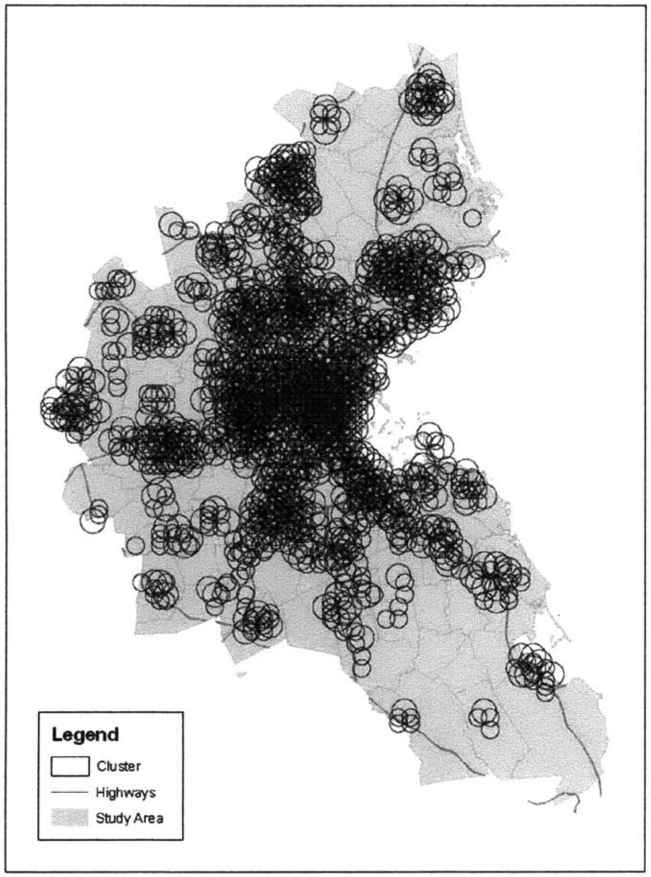


Figure 4-16 Clusters of Labor-intensive Firms at 2km Scale

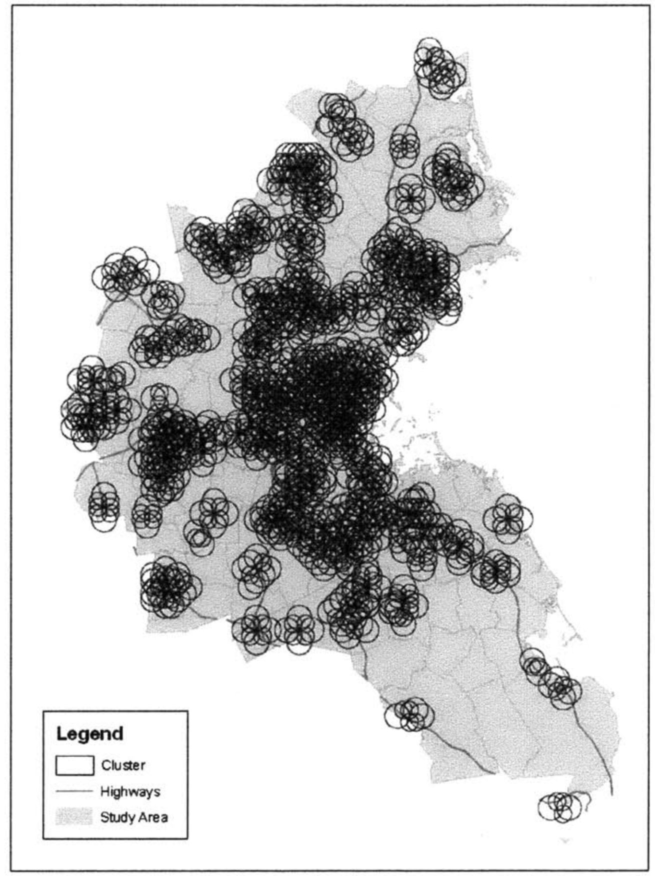


Figure 4-17 Clusters of Knowledge-intensive Firms at 2km Scale

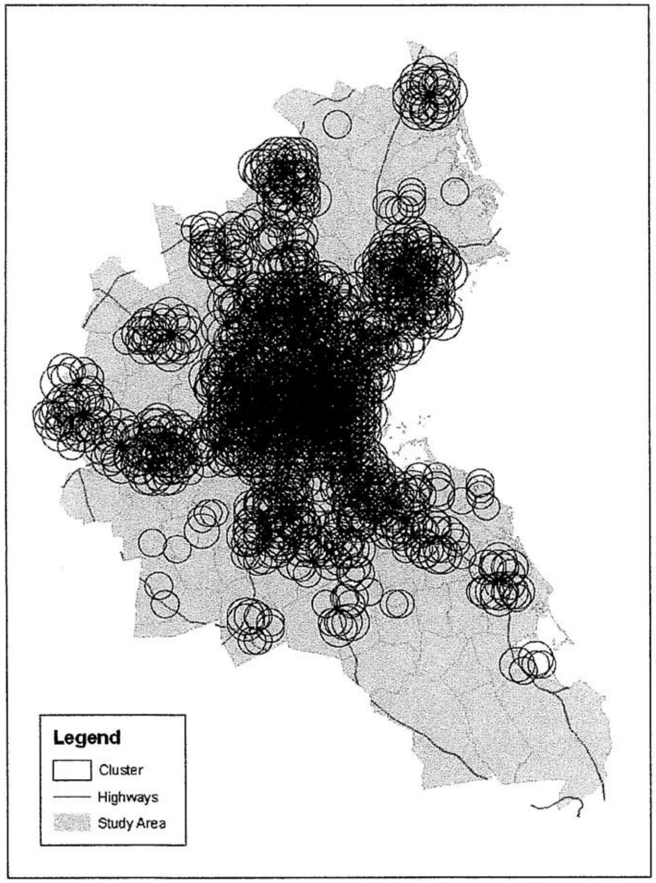


Figure 4-18 Clusters of Labor-intensive Firms at 3km Scale

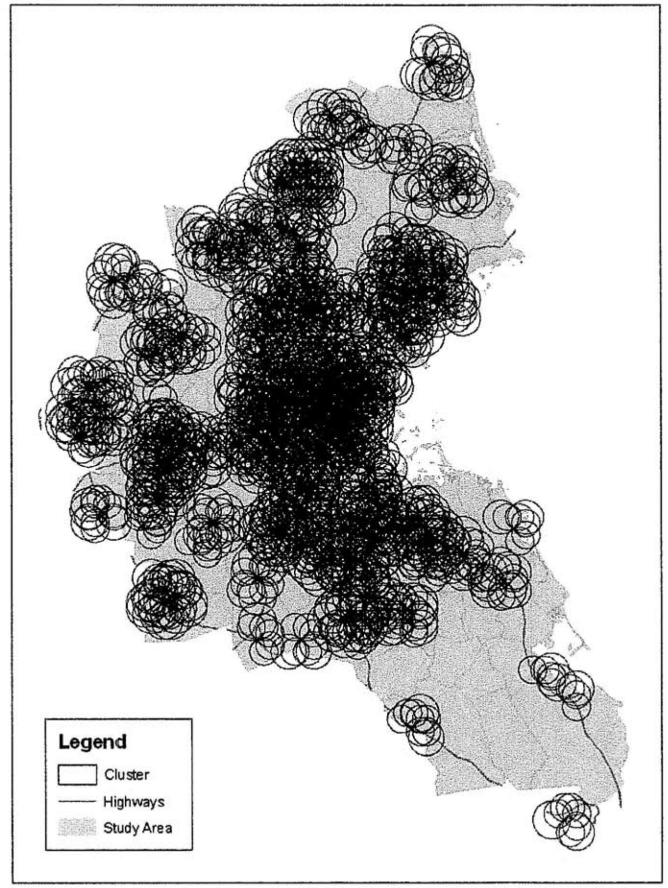


Figure 4-19 Clusters of Knowledge-intensive Firms at 3km Scale

4.4 Conclusion

In this chapter, with the help of GAM, we identify the significant industry clusters at .01 and .001 levels of significance, at scales between 1km and 3km. We compare the distributions of clusters of large plants with that of small plants to examine how plant size affects firm clustering; also, we compare the clusters of knowledge-intensive plants with that of labor-intensive plants to study how knowledge-dependency affects firm clustering. We observe ubiquitous existence of firm clustering in all firms, regardless of their size and technique differences. However, firms with different size and techniques are found with different clustering patterns. Although large firms are more clustered than small firms, their clusters are less location-restrained and scatter among many places along major highways. Knowledge-intensive plants are found to be less clustering than labor-intensive plants, and their clusters at small scales are found in fewer locations than that of labor-intensive ones.

From the result maps, two observations are of special interest: first, the strong relationships of major highways with clusters of urbanization: almost all clusters are found with one or more major highways adjacent; second, the super significant position of the downtown area of Boston for industry clusters: a large portion of the identified clusters of urbanization are found to locate in this area.

CHAPTER 5: DETECTING CLUSTERS OF LOCALIZATION

5.1. Introduction

In Chapter 4, we identified the significant clusters of urbanization at various scales scales. A noticeable feature of the results is the dominance of firm clusters are found in downtown Boston. This can be explained with the uneven distribution of economic activities, which exist in higher densities at city centers. In many circumstances, those locations with significant concentration of firms from specific industries are of more interest to researchers, since it is believed that clusters of localization are positively associated with firms' performance and regions' economic development, as proven in Silicon Valley's semiconductor industry, Hollywood's movie business, and the pharmaceutical industry in New Jersey and Massachusetts.

In this chapter, we introduce the method of detecting clusters of localization and compensating for the unevenly distributed background population. We identify the clusters of localizations in the Boston metropolitan area and compare them with those clusters of urbanization with the use of manufacturing firm establishment data of 2009.

5.2. Methodology

We can find very limited literature devoted to detecting firm clusters of localization. However, Kulldorff (1997)'s study of detecting "hot spots" of diseases with the background population unevenly controlled for provides excellent references. Similar to the strategies adopted by Openshaw (1987) and Turnbull et al. (1990), Kulldorff proposed to apply a series of circular zones of various scales to scan the study region in order to detect significant industry clusters at all scales. As the exact distribution of the test statistic could not be determined, the Monte Carlo procedure is also used for the significance test, just as we did in the previous chapter. To avoid the multiple testing problem arising in such methods as GAM, we use the maximum likelihood ratio test statistic, which was first proposed by Loader (1991), instead of the count in GAM.

Following Kulldorff's model, the model to detect clusters of localization can be developed as follows:

Assume there is exactly one circular zone ***Z*** such that for all business establishments within the zone, the probability of being a manufacturing plant is ***p***, whereas for those outside the zone, the probability is ***q***. The null hypothesis here is:

H_0: p=q,

and the alternative hypothesis is:

H_1: $z \in Z$, p>q.

The null hypothesis signifies the complete spatial randomness with each business establishment having an equal chance to be a manufacturing plant.

A likelihood function is defined to derive the test statistic:

$$L(Z,p,q) = \frac{e^{-p*\mu(Z)-q*(\mu(G)-\mu(Z))}[p*\mu(Z)+q*(\mu(G)-\mu(Z))]^n}{n!} \times \prod_{x_i \in Z} \frac{p*\mu(x_i)}{p*\mu(Z)+q*(\mu(G)-\mu(Z))} \times \prod_{x \notin Z} \frac{q*\mu(x_i)}{p*\mu(Z)+q*(\mu(G)-\mu(Z))} \qquad (5\text{-}1)$$

where:

n: the total number of manufacturing establishments,

G: the study region,

$\mu(G)$: count of business establishment within G,

$\mu(Z)$: count of business establishment within Z,

$\mu(x)$: count of business establishment within location x.

Equation (5-1) defines the probability of observing zone ***Z***, and it consists of three parts: the first part is the probability of observing ***n*** manufacturing establishments within the study region; the second part is the probability of observing all establishments within the clusters, and the third part is the probability of observing all establishments outside of the clusters. Their product gives the likelihood of observing the existing establishment distribution within the study region as Equation (5-1), which can be further simplified as:

$$\frac{e^{-p*\mu(Z)-q*(\mu(G)-\mu(Z))}}{n!} \times p^n q^{N-n} \times \prod_{x_i} \mu(x_i), \quad (5\text{-}2)$$

where:

N: number of business establishments within the study region.

A test statistic is defined as:

$$\lambda = \frac{\sup_{z\in Z, p>q} L(Z,p,q)}{\sup_{p=q} L(Z,p,q)} \qquad (5\text{-}3)$$

This test statistic defines the most likely cluster in the data, which has the maximum deviation from the randomness. To calculate λ, it is essential to estimate Z, which meets Equation (5-3).

As the denominator in Equation (5-3) can be simplified as:

$$\sup_{p\in[0,1]} \frac{e^{-p*\mu(G)}p^n}{n!}\prod_{x_i}\mu(x_i) = \frac{e^{-n}}{n!}(\frac{n}{N})^n\prod_{x_i}\mu(x_i) = L_0 \qquad (5\text{-}4)$$

The numerator in Equation (5-3) is maximized by two steps: first the maximum is taken over all p and q for a fixed z. Equation (5-3) attains the maximum when: *$p=n_z/\mu(Z)$ and $q=(n-n_z)/(N-\mu(Z))$*, and the value is:

$$L(Z)=\begin{cases} \dfrac{e^{-n}}{n!}\times\left(\dfrac{n_z}{\mu(z)}\right)^{n_z}\times\left(\dfrac{n-n_z}{N-\mu(z)}\right)^{N-n_z} & \text{if } \dfrac{n_z}{\mu(z)} > \dfrac{(n-n_z)}{(\mu(G)-\mu(Z))} \\ \dfrac{e^{-n}}{n!}\times\left(\dfrac{n_z}{\mu(z)}\right)^{n_z}\prod_{x_i}\mu(x_i) & \text{otherwise} \end{cases}$$

Thus, we have

$$\lambda = \frac{\max_z L(z)}{L_0} = \sup_{z\in Z}\frac{(\frac{n_z}{\mu(z)})^{n_z}(\frac{n-n_z}{N-\mu(z)})^{n-n_z}}{(\frac{n}{N})^n} I(\frac{n_z}{\mu(z)} > \frac{(n-n_z)}{(\mu(G)-\mu(Z))}) \qquad (5\text{-}5)$$

In Equation (5-5), the I() is an indicator function. This likelihood ratio is maximized over all circular zones, and the zone with the maximum value of λ is identified as the most likely industry cluster. In some cases, as in this study, we also identify the second, third,... most likely industry clusters, which have the second, third, ... largest value of λ.

5.3. Computation and Inference

The computation of λ is performed with the help of SaTScan V7.02, which is specially developed to implement the likelihood ratio test by Kulldorff, et al. (2007). Three files are provided for the calculation in SaTScan: coordinate file, population file and case file. The coordinate file contains the geographic coordinates for the locations. In our study, a grid of $500x500m^2$ is laid over the study region, and we regard the centroid of each cell as a location. The population file provides information about the background distribution of all business establishments. For each location, the population is the count of business establishments. The case file defines the information about manufacturing establishments, and the count of manufacturing establishments is input for each location.

To correct the "Edge Effects", we adjust the count of establishments is adjusted according to the ratio of each circular zone's portion within the study region and that outside of the study region. For example, if ***m*** establishments are found within a zone, which have ***n%*** within the study region, then we adjust the count to be ***m/(n%)***. The adjustment is only carried out for land adjacent areas, and those coastal areas are not adjusted, since there actually is nothing in the ocean.

As the likelihood ratio test statistic depends on the underlying nonhomogeneous distribution of business establishments, it usually has no simple analytical form. To determine the significance of the calculated value, we conduct a Monte Carlo simulation. In the Monte Carlo simulation, the calculation of the likelihood ratio λ is repeated over a large of random (999 in this study) replication under the null hypothesis and the p value of λ is determined according to its rank among the 1000 values of λ (999 samples + 1 observation). The test of the hypothesis is also carried out through SaTScan.

By using the spatial scan statistic, while identifying the most likely cluster within the study area, we can also identify and infer those secondary "clusters". However, as Kulldorff pointed out, it will give a conservative result when comparing the secondary λ with those most likely estimates from the random samples. And it will also be unsatisfactory to compare the secondary λ with those secondary most likely estimates from the random simulations, since "it would not account for the size of the primary cluster in the data" (1994).

5.4. Detecting Clusters of Localization in Boston

5.4.1. Data

The data used in this study are the same as that in Chapter 3, including the firm establishment data from InfoUSA, 2009, and the education attainment data for industries comes from Current Population Survey (CPS) 2006, provided by Bureau of Labor Statistics (BLS).

5.4.2. Results

a) Significant clusters of localization within Boston Metro

Figure 5-1 and 5-2 present the existing distributions of two types of firms: manufacturing firms and all firms. The distribution of all firms also defines the economic backgrounds of manufacturing firm clustering. These two distributions exhibit some similar features, for example both seem to follow a hub-and-spoke structure with downtown as the center, some areas as hubs and major highways as spokes that connect hubs. Of course, they also have differences. Those clustering features that reside in the distribution of manufacturing firms but not in the distribution of all firms represent some special clustering preferences favored by manufacturing firms. To understand these special clustering features of manufacturing firms, it is important to identify those areas with exclusive concentration of manufacturing firms. These are the localization clusters depicted in from Figure 5-3 to 5-8. In these maps, two types of localization clusters are listed side by side for comparison: localization clusters of establishment and localization clusters of employment. From the maps, we can see that contrary to urbanization clusters' predominantly locating within downtown areas, localization clusters mainly appear in suburban areas and stretch along highways or major roads.

At 1km scale, localization clusters of establishment scatter around almost the whole study area with two important hubs: Woburn areas in the north of Boston and areas around Burlington, Norwood and Randolph in the south of Boston. At 2km scale, localization clusters of establishment become even more compactly distributed and are found closely stretching along route 128. At 3km scale, localization clusters of establishment become dispersed again and scatter along highway 495. From here we also see that the pattern of firm clusters does not always vary in same direction when the spatial scale changes, thus findings at one spatial scale cannot be simply applied to other spatial scales.

Compared with localization clusters of establishment, localization clusters of employment follow similar pattern with the growth of spatial scales. Along three major highways: highway 495, I93 and route 128, firm clusters become more dispersed from 1km scale to 3km scale.

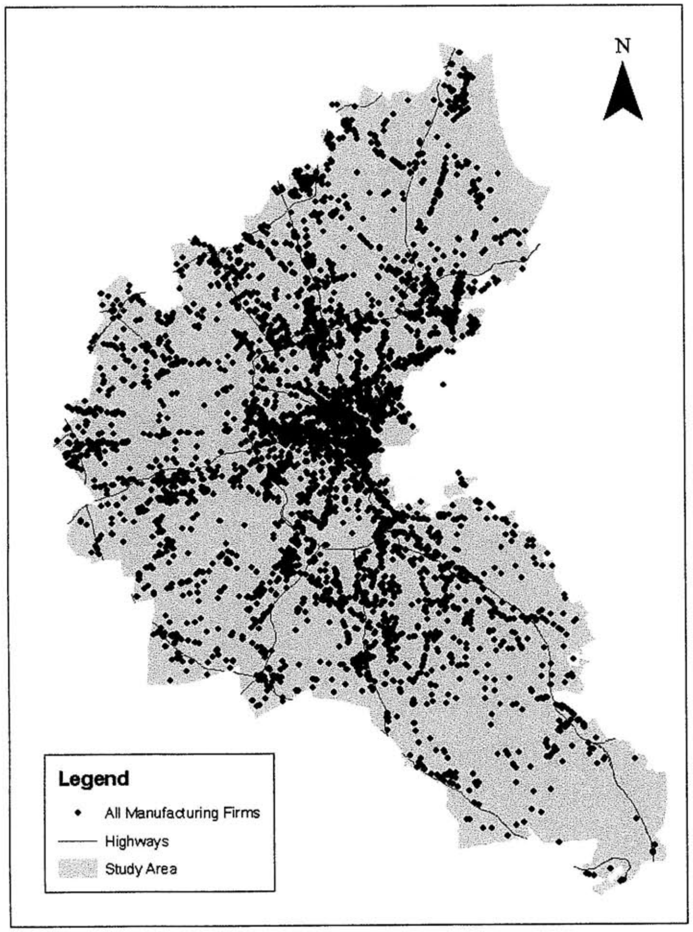


Figure 5-1 Map of Manufacturing Plants

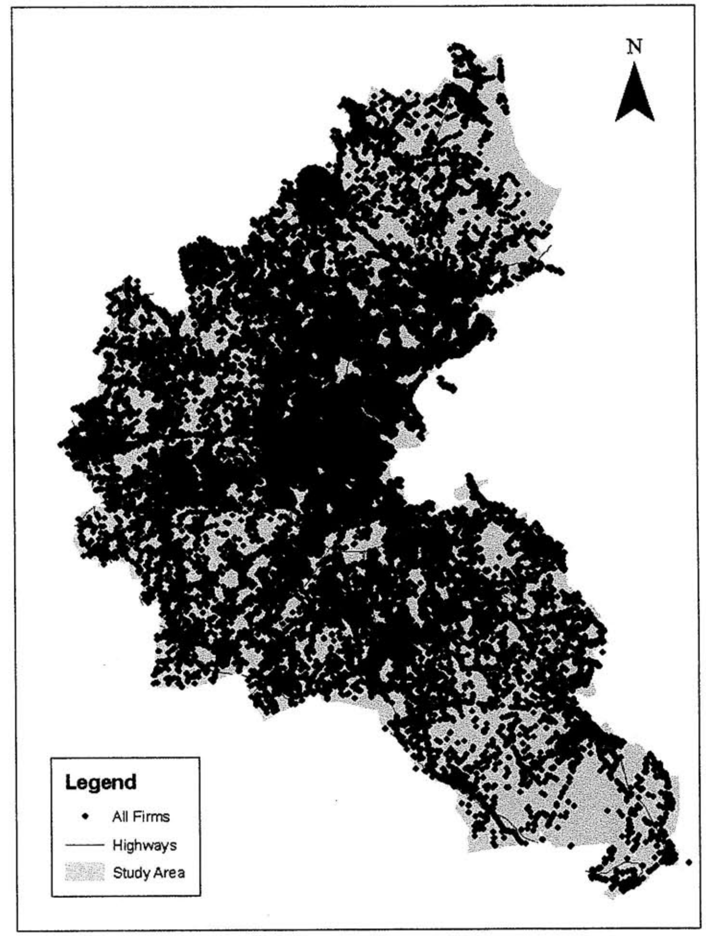


Figure 5-2 Map of All Business Establishments

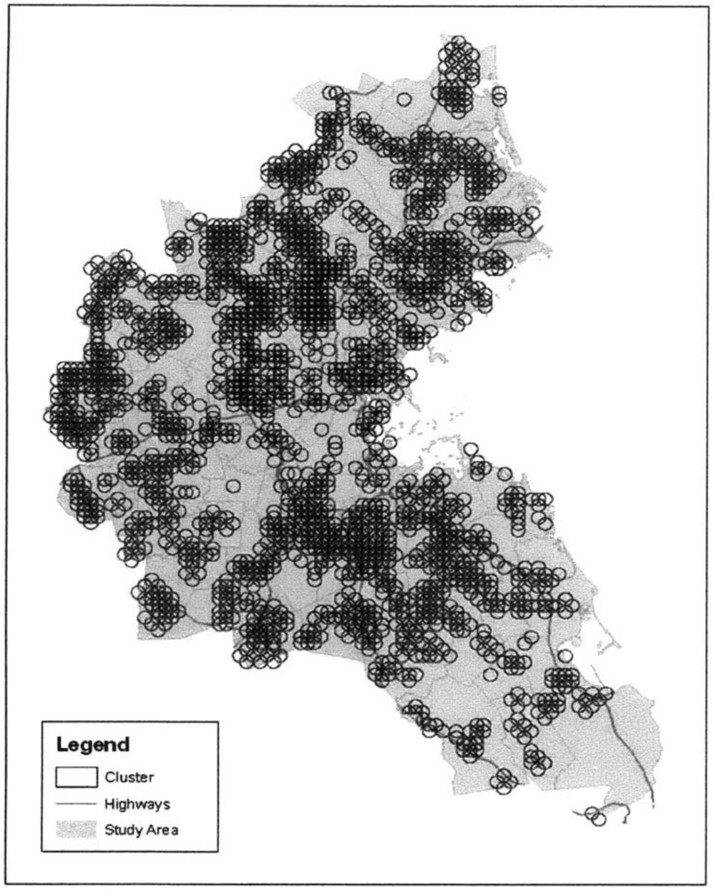


Figure 5-3 Localization Clusters of Establishment, 1km

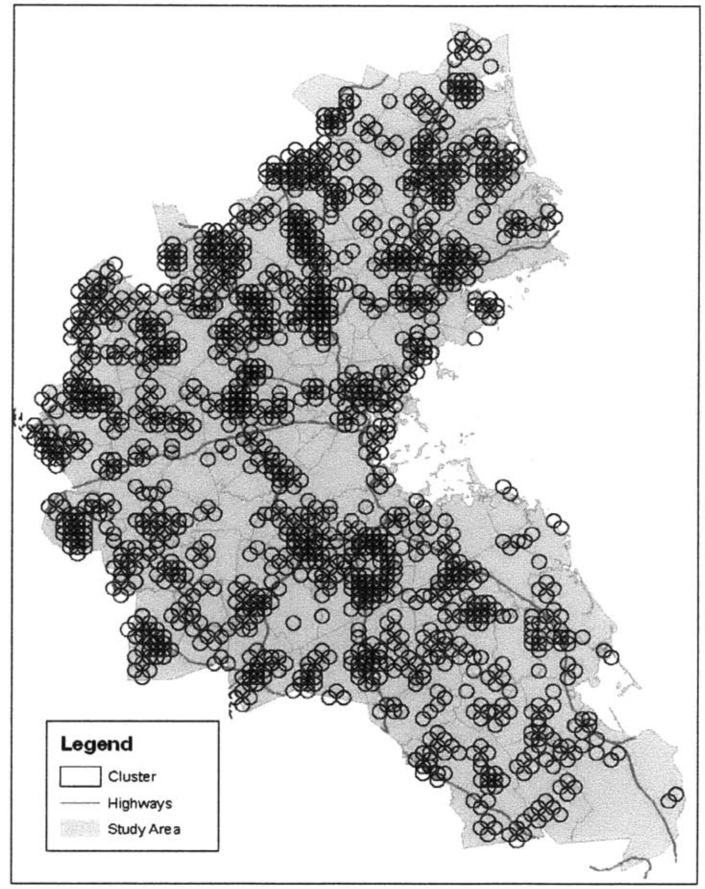


Figure 5-4 Localization Clusters of Employment,1km

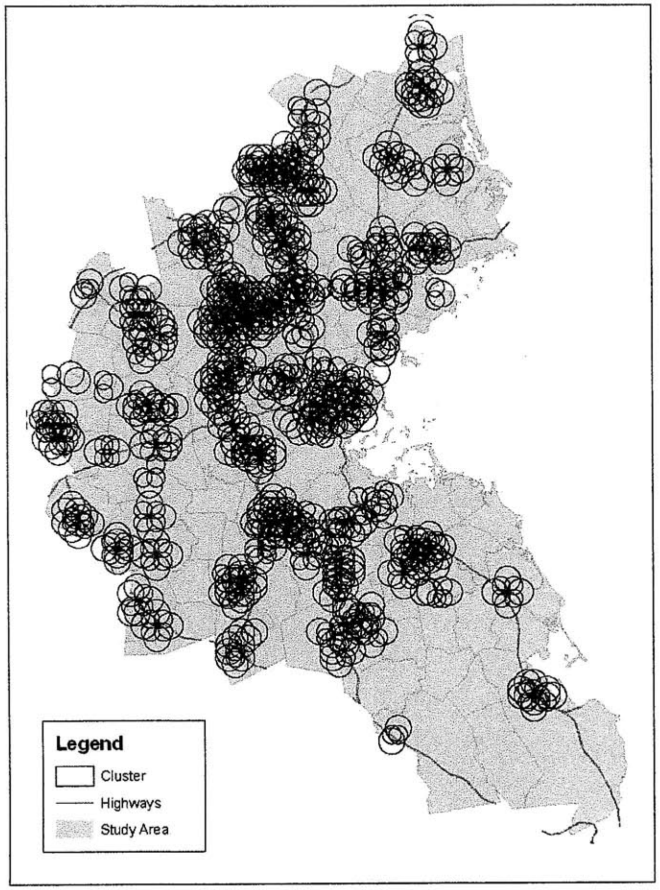


Figure 5-5 Localization Clusters of Establishment, 1-2km

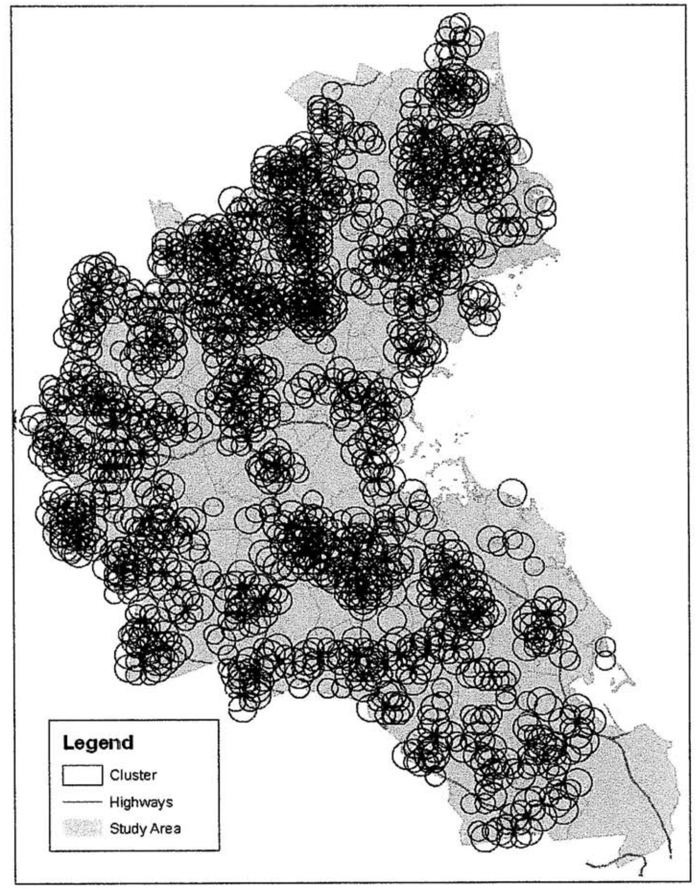


Figure 5-6 Localization Clusters of Employment, 1-2km

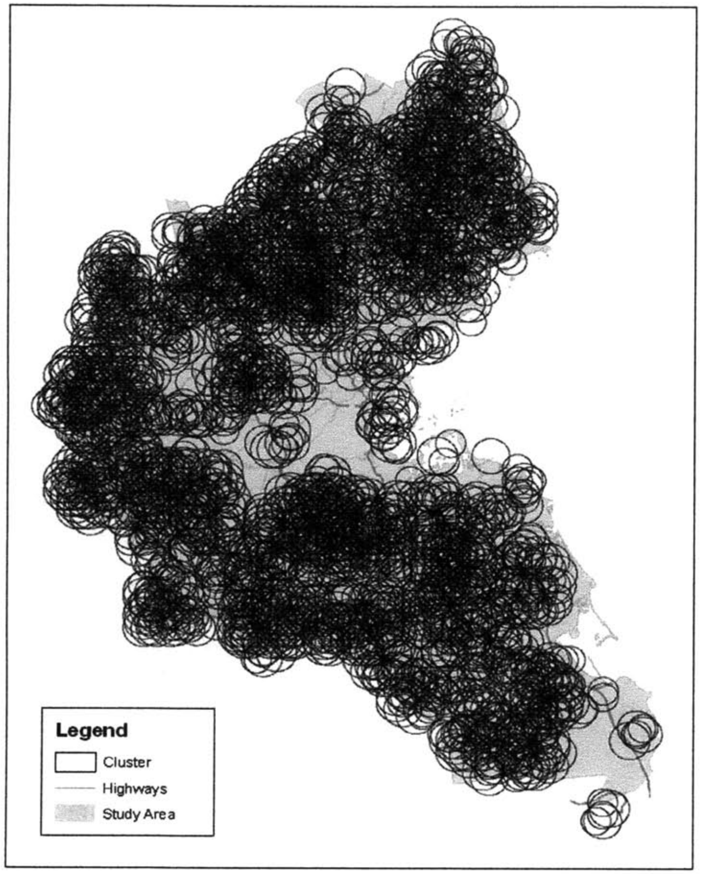


Figure 5-7 Localization Clusters of Establishment, 2-3km

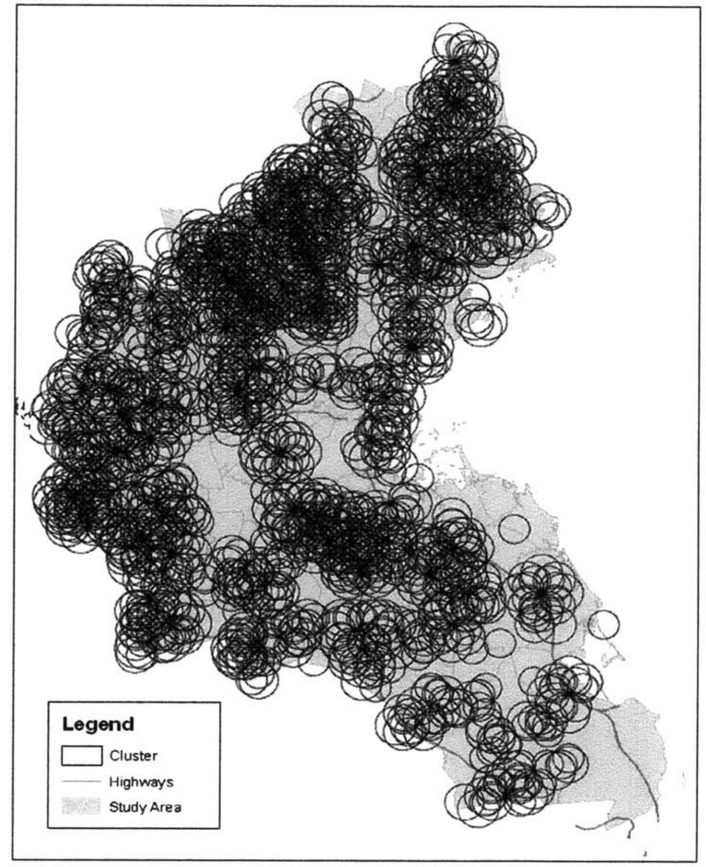


Figure 5-8 Localization Clusters of Employment, 2-3km

b) Cluster of localization and the plant size

In this session, localization clusters of large firms are compared with those of small firms to observe the impacts of plant size on firm clustering. The definition of large firms is consistent with that used in Chapter 3, and the cut-off value for large size firms is also set to be 35. This gives us 2056 large firms, which account for 26% of the total manufacturing establishments.

Figures 5-9 to 5-14 display the significant large firm clusters and small firm clusters identified at scales between 1km and 3km, at the .001 level of significance. A major difference between the clusters of large firms and those of small firms disclosed by these maps is the large firms' compact pattern vs. the small firms' dispersed pattern. At 1km and 2km scales, clusters of large firms appear in a very compact style: mainly along route 128 and in the downtown area. On the contrary, clusters of small firms appear in much wider areas. At 3km level, clusters of large firms become a little more dispersed and a group of clusters scatter along highway 495; clusters of small firms become more dispersed with the increase of spatial scales.

If we compare findings we have here with what we found at chapter 4 (Figure 4-8 to 4-13), we can find that urbanization clusters and localization clusters are highly consistent in their patterns regarding plant size: large firms tightly scatter along highway corridors and small firms disperse across the study region. As large manufacturing firms normally demand large lands and office space, they turn to those areas that other economic activities pay little attention to, like highway corridors, while their noise, separation and congestion keep residential or commercial investors away, and their space and highway accessibility become manufacturing firms' ideal clustering locations.

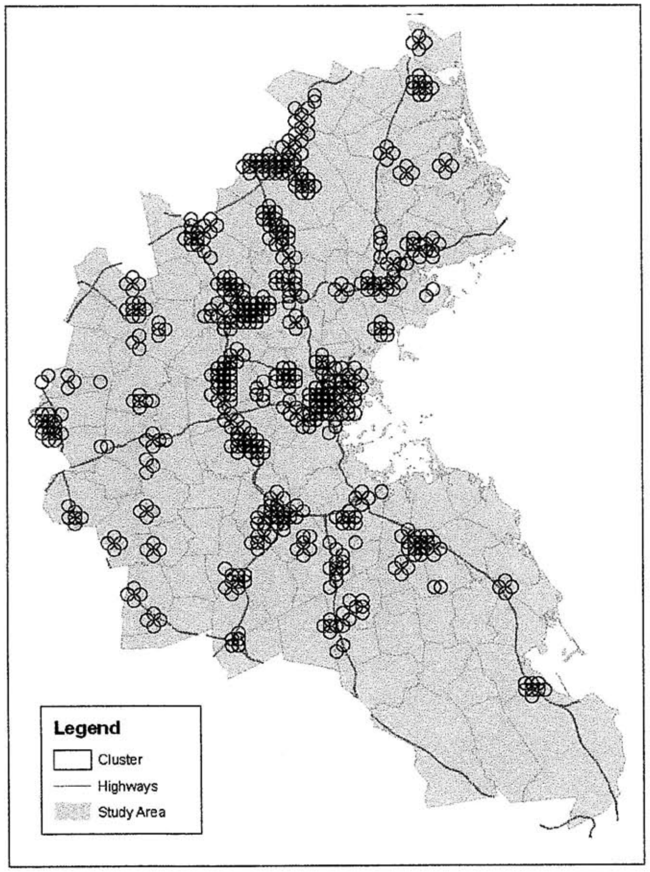


Figure 5-9 Localization Clusters of Large Firms, 1km

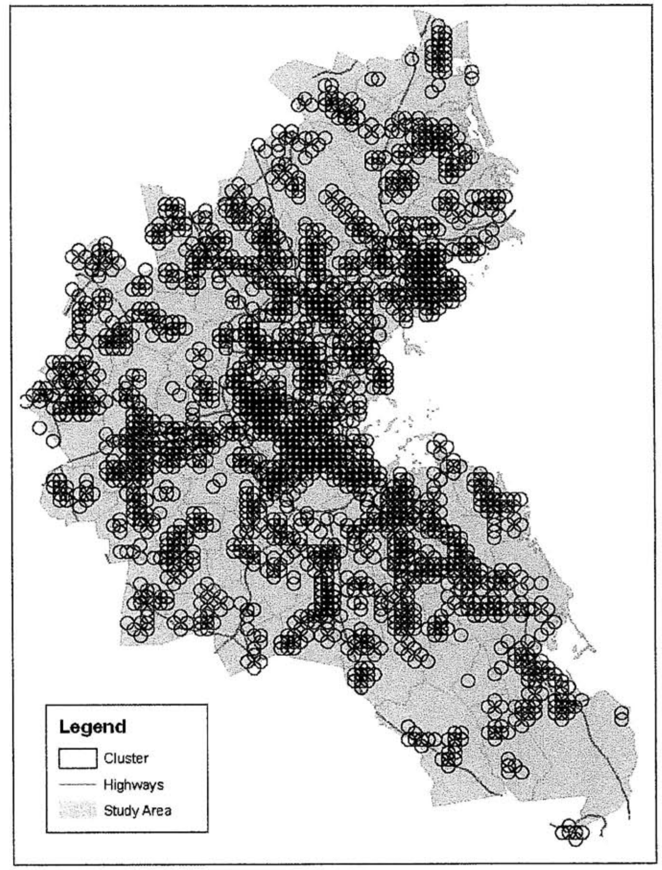


Figure 5-10 Localization Clusters of Small Firms, 1km

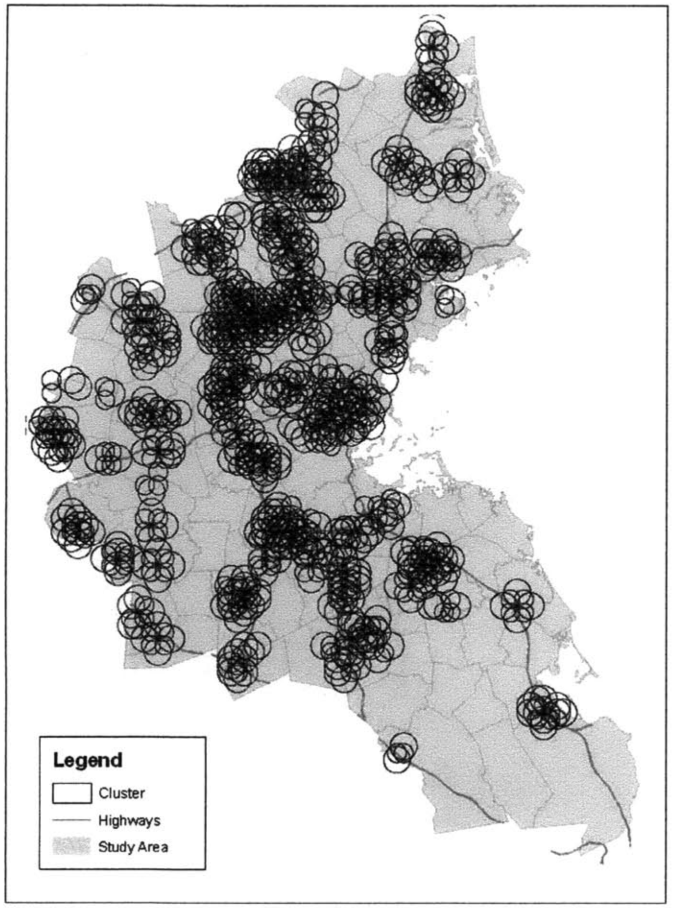


Figure 5-11 Localization Clusters of Large Firms, 1-2km

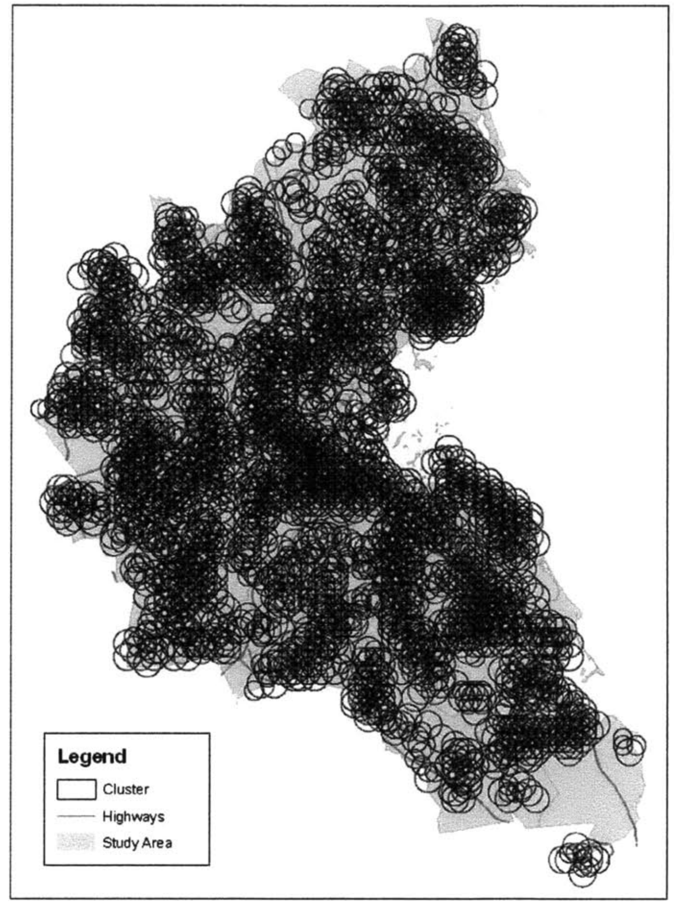


Figure 5-12 Localization Clusters of Small Firms, 1km

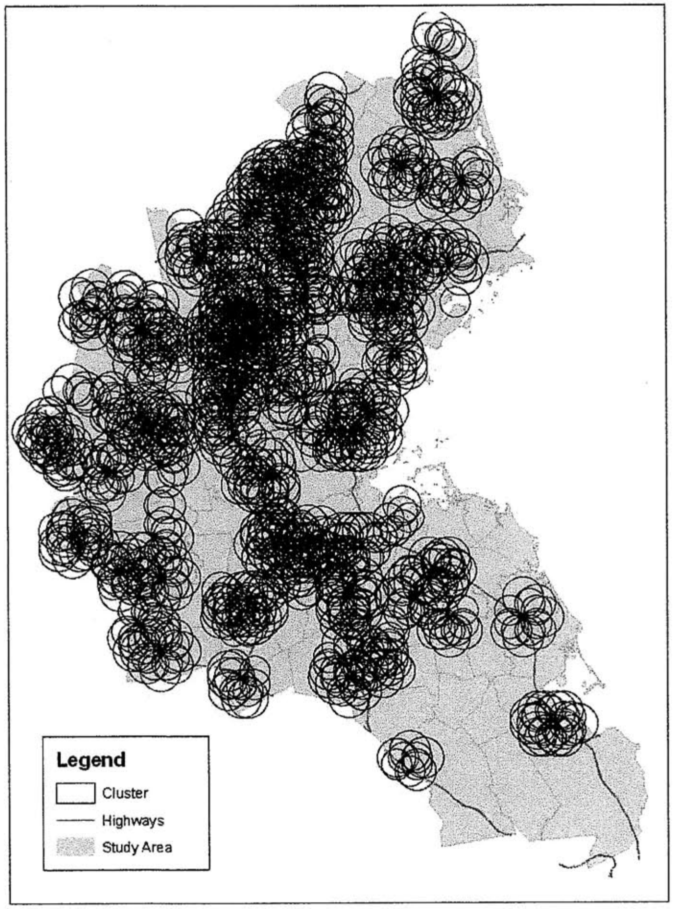


Figure 5-13 Localization Clusters of Large Firms, 2-3km

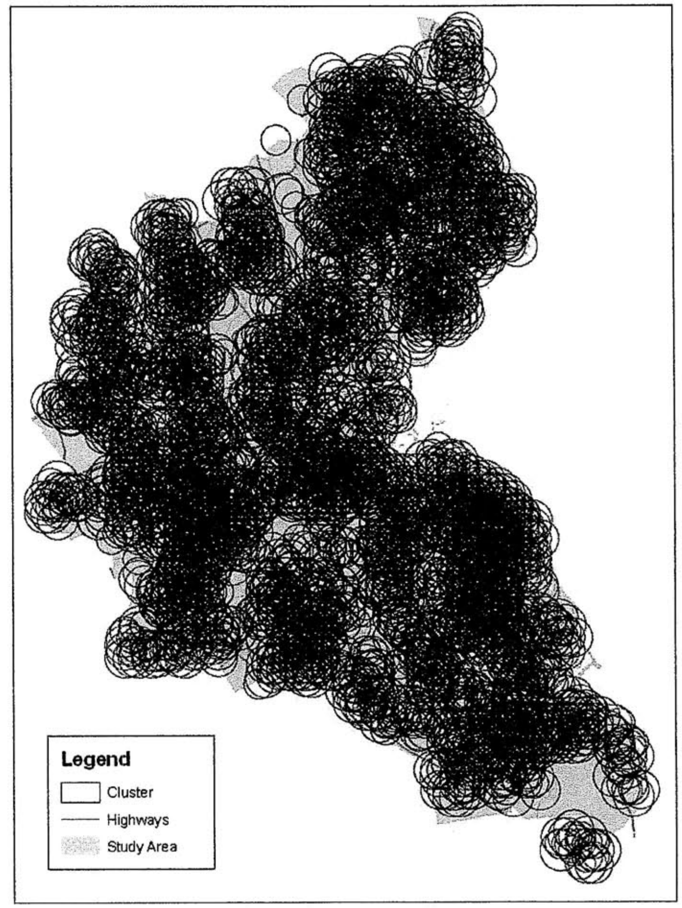


Figure 5-14 Localization Clusters of Small Firms, 2-3km

c) Firm's knowledge intensity and clusters of localization

To find out how technical intensity affects firm clustering, we divide all plants into two groups: knowledge-intensive plants and labor-intensive plants as we did in Chapter 3. This gives us 2029 plants identified as knowledge-intensive firms, and 5746 as labor-intensive firms. For more information about this partition, please reference Table 4-2.

Figures 5-15 to 5-20 show the distribution of localization clusters of knowledge- and labor-intensive plants at scales between 1km and 3km. Compared with our findings about the impacts of technical intensity on urbanization clustering in chapter 4 (shown as 4-14 to 4-19), we observe very similar patters of firm clusters: at 1km scale, both types of clusters almost duplicate their corresponding distributions in previous chapter; at 2km scale, some labor-intensive clusters appear in the south of study area around Bridgewater town and along route 3, but knowledge-intensity clusters of localization do not vary too much from their urbanization correspondence; at 3km scale, both types of clusters just strengthen their patterns shown as 2km scale.

As we discussed in chapter 4, researchers find that technical-intensive firms are less concentrated than labor-intensive firms and we find that this is not the fact at detailed scale for urbanization clusters. Here we find that it is true for localization clusters at almost all three scales too. However, as the pattern of firm clusters does not equal to the pattern of firms and our findings are based on one case study, it is still possible that findings in literature still hold for some other regions.

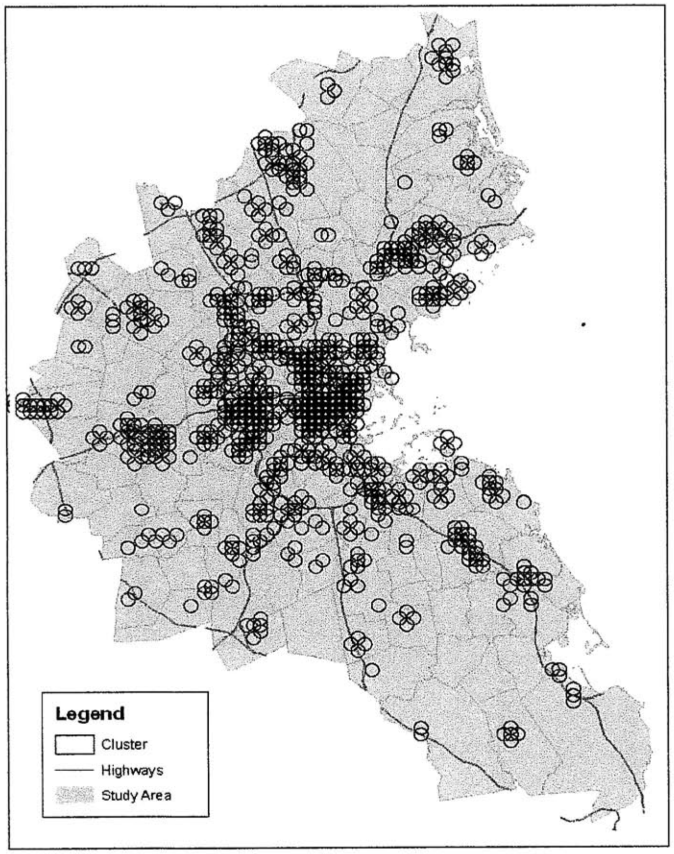


Figure 5-15 Localization Clusters of Labor-intensive Firms, 1km

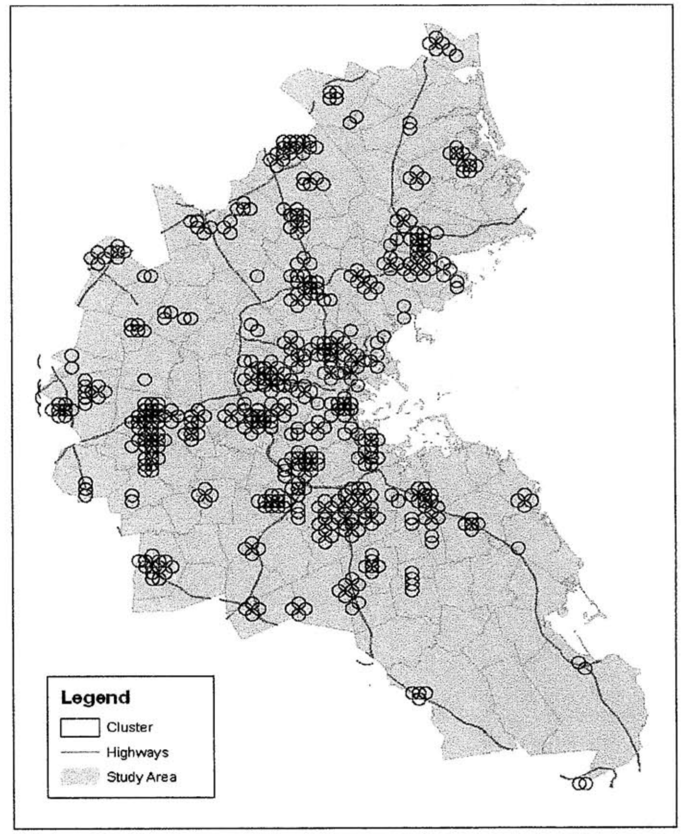


Figure 5-16 Localization Clusters of Knowledge-intensive Firms, 1km

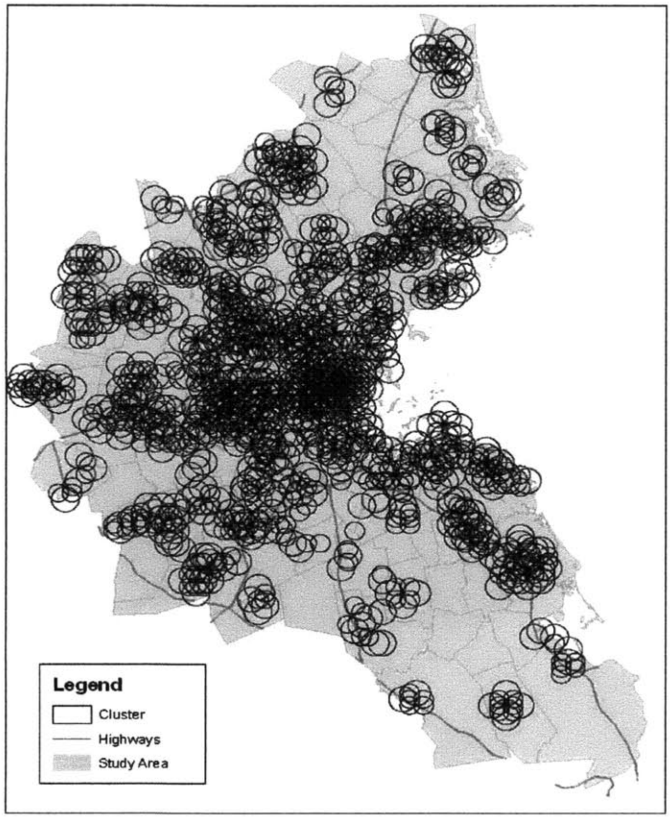


Figure 5-17 Localization Clusters of Labor-intensive Firms, 1-2km

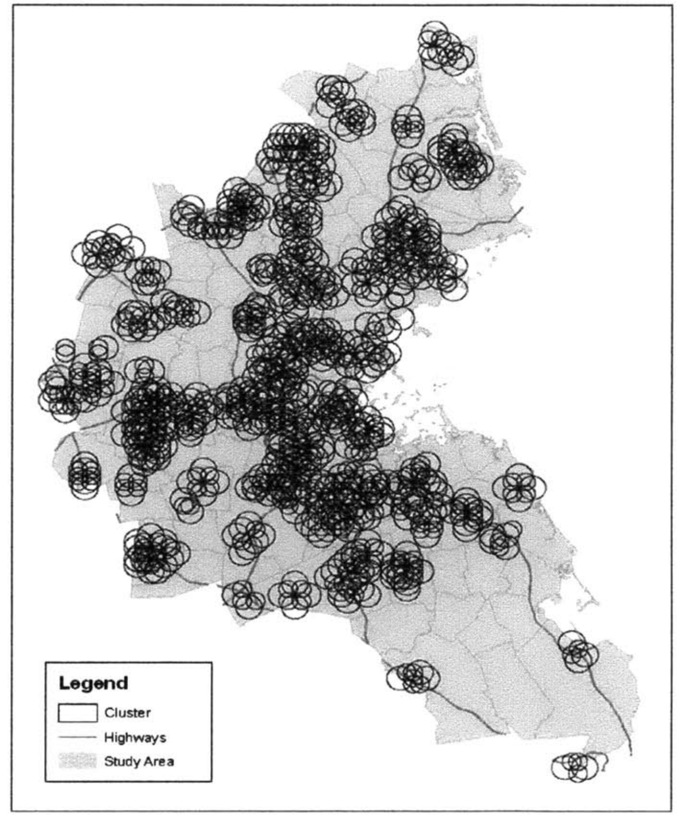


Figure 5-18 Localization Clusters of Knowledge-intensive Firms, 1-2km

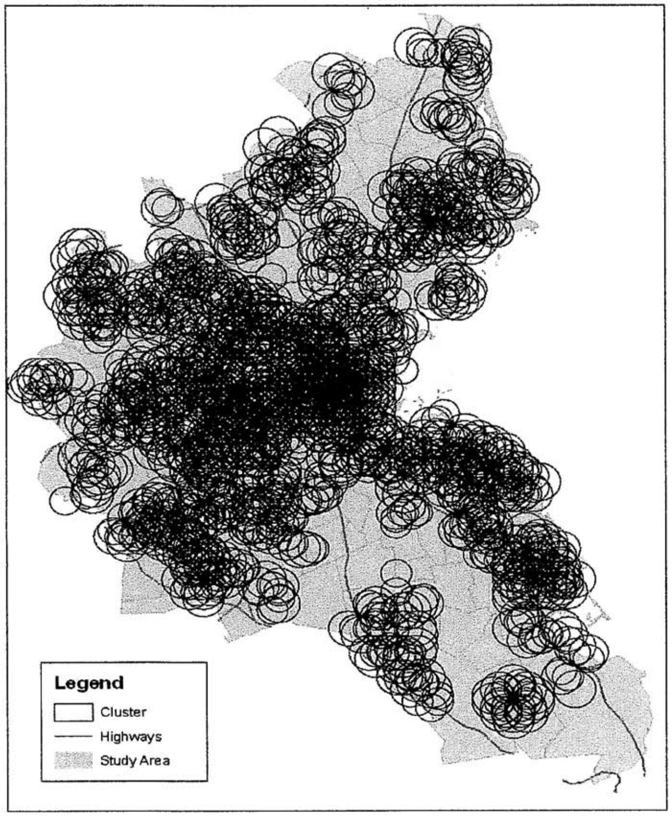


Figure 5-19 Localization Clusters of Labor-intensive Firms, 2-3km

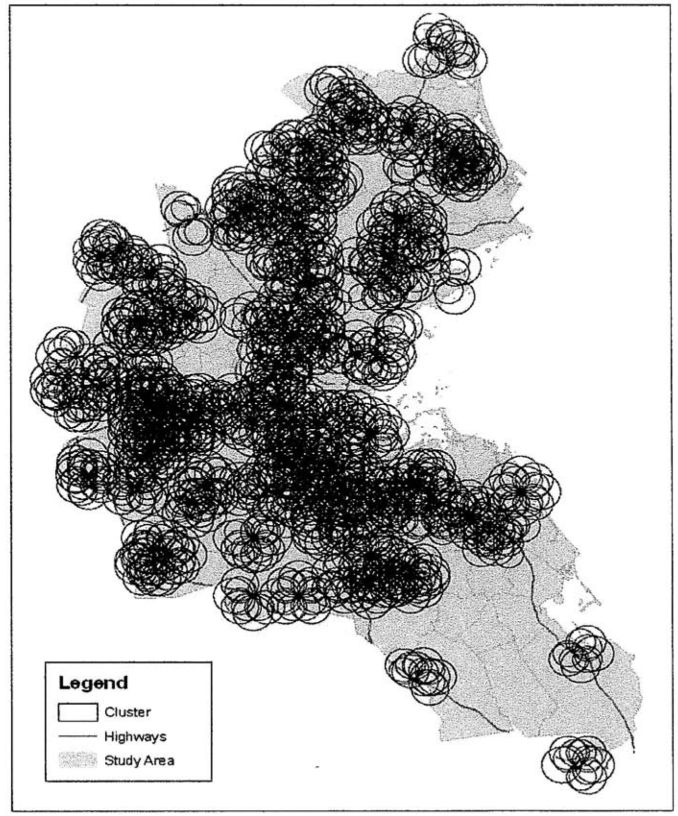


Figure 5-20 Localization Clusters of Knowledge-intensive Firms, 2-3km

5.5 Conclusion

In this chapter, we introduce a term: localization clusters and the methodology to identify them. While urbanization clusters refer to the disproportional concentration of manufacturing firms compared with spatial randomness, localization clusters refer to the disproportional concentration of manufacturing firms compared with other firms. By empirically detecting the clusters of localization and comparing it with our findings in Chapter 3, we can see that clusters of localization and clusters of urbanization have obviously different patterns: while clusters of urbanization are dominantly found within the downtown area, the clusters of localization mainly distribute within the close suburban areas along major highways such as Route128 and I-90 West. Similar to our study in Chapter 3, we specifically look into how firm size and knowledge-dependency affect firm clustering. We find that large plants have fewer and more separated clusters at small scales than small plants, but they have similar patterns at large scales. However, in our study of clusters of urbanization in Chapter 3, large plants are found to have more clusters that are smaller in size and more scattered in space. We also find that knowledge-intensive plants have obviously dissimilar distribution from those of labor-intensive plants. While labor-intensive plants cluster mainly along Route 128 and I-90 West, knowledge-intensive plants are found to be more natural-amenity oriented.

Although we have the uneven distribution of background business establishments controlled for, two significant factors that are found to affect clusters of urbanization in Chapter 3, major highways and downtown, are re-verified by the study of clusters of localization. However, instead of having large numbers of clusters of urbanization located within downtown, dominant clusters of localization are distributed around downtown area along Route 128; and instead of stretching along several major highways in all directions like clusters of urbanization, clusters of localization are mainly found along highways around downtown such as: I-495 and I-90 West.

CHAPTER 6: DETERMINANTS OF FIRM CLUSTERING

6.1. Introduction

In Chapters 4 and 5, we introduced the firm cluster detection methods and looked into the distribution of firm clusters within metropolitan Boston. In these studies, we observed the ubiquitous existence of firm clusters at all spatial levels and learned that firms with different employee numbers and technical intensities can expose different clustering patterns. This chapter continues our study by investigating the determinants of firm clustering. Why do some firms locate within clusters, and some do not? Do firms' characteristics like size, productivity and age make a difference in firms' clustering decisions? As noted in previous chapters, while there are extensive theories and empirical evidence about agglomeration economies and industrial clustering, little empirical research has focused on firm clustering at the intrametropolitan level and none has been found systematically to detect firm clusters within metropolitan areas in terms of location and scale. Without identified clusters, those questions concerning linkages between firms and clusters, and between geographic locations and clusters, are mostly left untouched.

It is important to note that findings at a regional level are not guaranteed to also hold at the firm level, as studies have shown that relationships regarding firm clustering change quickly at a micro level. For example, Rosenthal and Strange (2003), Cainelli and Lupi (2008), and Duranton and Overmann (2005) employed different approaches to measure the changes of agglomeration economies with respect to the geographic distance, and find that agglomeration effects are strongest in the nearest vicinity and tend to attenuate quickly. When data are aggregated to a regional level, those important dynamic variations in firm clustering often get lost. Moreover, findings about clustering and agglomeration at the regional level cannot replace research results based on firm clusters. With clearly defined boundaries and locations, firm clusters can tell us about firms' clustering behaviors as well as their location preferences, which provides strong policy implications to urban planners in making effective land-use arrangements.

This chapter presents a humble attempt to reduce these gaps in the literature regarding firm clustering at the intrametropolitan level by explicitly linking the detected firm clusters to the firm's characteristics. The objective in this chapter is to study how firms' characteristics associate with their clustering behaviors. To do this, we use a logistic regression model of firm's clustering decisions that is specific to the independent characteristics of the firm. We calibrate the model using a manufacturing firm establishment data set within the Boston metropolitan area.

This chapter is organized as follows: section 2 provides a brief literature review on the relationship between firm's characteristics and firm's clustering behaviors; section 3 introduces the econometric model and estimation issues; and section 4 presents the results and discussion.

6.2. Literature review

Numerous studies have identified a wide range of firm characteristics that are claimed to affect firm clustering behaviors in one way or another. For example, Wolf and Pett (2000) found that firm size plays an important role in firm's clustering behavior. They argued that as large firms have the capacity to internalize many of the advantages of clustering (labor talent, access to markets, business services), they are more reluctant to locate within firm clusters. Small firms, on the contrary, are willing to endure the diseconomies of urbanization to gain otherwise unavailable external scale economies by locating within firm clusters. If this holds true at the firm level, we should observe negative impacts on firm size upon choosing to locate within firm clusters.

As new firms are often small in size, based on Wolf and Peltt's theory, new firms should be more likely to locate within firm clusters. This is tested by Phelp, Fallon and Williams (2001) and they found that new firms try to locate in close spatial proximity of the existing clusters in order to utilize economic externalities, and with well established firms, firm clusters can create an infrastructure suitable for the new firms. However, not all findings are consistent. Yuji Honjo (2000) argued that firm clustering causes negative impacts on the survival of new firms because of the occurrence of higher competition in the center, and larger firms' uncooperative attitudes can also hinder small firms' enthusiasm in locating within clusters (Okamuro, 2008).

Most literature agrees that technical spillover is one of the important factors that cause firm clustering and agglomeration. Alfred Marshall and Paul Krugman considered technical spillover along with labor market pooling and specialized inputs as three major reasons for agglomeration. It is believed that geographic proximity can facilitate the formal and informal interactions among workers, which can generate positive externalities for firms. Saxenian (1994) argued that the enormous success of Silicon Valley compared to Massachusetts' Route 128 lies in the former's geographic proximity of high-tech firms and the social network among them, which facilitated a more efficient flow of new ideas and ultimately caused faster innovation. Jaffe et al.

(1993) also found patent citations are more likely to come from the same state or metropolitan area as the originating patent. While it is unclear whether or not these theories hold at the firm level or not and in what form if they do, we should observe a closer association between a firm's technical dependencies and its clustering behaviors as technically-intensive firms are more sensitive to technical spillovers, which is strongest in firm clusters.

Researchers also find a firm's organizational structure imposes impacts on the firm's clustering behavior. DeYoung, Robert et al. (2004) observed that while bank headquarters move to those larger cities with agglomeration economies available to banking companies, bank branches have moved substantially farther away from bank headquarters. Hong, Junjie (2007) also found that compared with independent firms, branch companies are less responsive to local market demand and labor factors, since for branch companies, scale economies and knowledge transmission are mostly internalized within the company.

6.3 Methodological issues

6.3.1. The econometric model

Consider a manufacturing firm that has decided to open its factory in a metropolitan area. The site options are those locations assigned with industrial land uses by zoning, which are distributed on a discrete urban space. A firm will choose a location that brings it the highest utility level. In terms of clustering, each firm's final location decision will be one of two possibilities: within or outside a cluster. Our hypothesis is that a firm's clustering decision, whether to locate within a cluster of a certain scale, is not a random occurrence and it can be explained by the firm's specific characteristics. In this circumstance, traditional ordinary least squares (OLS) regression is not appropriate any more since the dependent variable, firm's clustering decision, can only be one of two values: one or zero, which is not a continuous variable and cannot be negative. The logistic regression model can be employed when the dependent variable is zero or one and it has a form as follows:

$$P(y_i = 1) = \frac{1}{1 + \exp^{-f(u)}},$$

or $$\log it(P) = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = f(u), \quad (6\text{-}1)$$

where $P(y_i = 1)$ is the probability that a firm chooses to locate within a cluster of a specified scale, u is the firm's utility level that is specific to the firm's independent characteristics and explains the firm's clustering decision. In literature, U is estimated with different forms. Following Carlton (1983) we are going to use a modified Cobb-Douglas production function to model the utility level that a firm can reach by locating within clusters of certain scales:

$$u = K_0 X_1^{\alpha_1} X_2^{\alpha_2} \ldots X_m^{\alpha_m}, \quad (6\text{-}2)$$

where K, $\alpha_1 \ldots \alpha_m$ are unknown constants and $X_1 \ldots Xm$ are firmspecific variables.

For estimation convenience, we use a logarithm form of (6-2), as follows:

$$\ln u = \ln K_0 + \alpha_1 \ln X_1 + \alpha_2 \ln X_2 + \ldots + \alpha_m \ln X_m \quad (6\text{-}3)$$

Combining (6-1) and (6-3), we have the general model that we use in this chapter:

$$\log it(P) = \ln K_0 + \alpha_1 \ln X_1 + \alpha_2 \ln X_2 + \ldots + \alpha_m \ln X_m \qquad (6\text{-}4)$$

6.3.2. Selection of study clusters

In metropolitan Boston, the detected firm clusters consist of a hierarchical system of clusters with different locations, sizes and extent. In geography they overlay each other. In size, they range from street corner stores in the city of Boston to the whole Boston metropolitan area. As clustering is a natural feature of human activities and the available industrial land uses are limited to a few locations, all manufacturing firms in metropolitan Boston locate within clusters, one way or another. The only thing that makes a difference is the scale of clusters in which they locate in: it is a block level cluster or it is a city-level cluster. To make the study meaningful, selections of clusters of certain scales are necessary, since it does not make much sense to study why firms within metropolitan Boston locate within a metropolitan level cluster, which is the study area itself. The selection criteria depend on the research interest and the size of the study area. For our study, we are interested in understanding firm clustering behaviors at an intrametropolitan level and comparing them with available regional-level findings, which can be found in literature. Thus we select those identified firm clusters that are smaller than 1.75km in radius. We use 1.75km mostly because it is the average size of townships within the study area. Figure 6-1 shows the final selection of firm clusters within the study area.

Because of the overlapping nature of firm clusters, to avoid the redundancy of data, we also require that for any firm cluster, the overlapping area cannot exceed 50% of its area, otherwise, only the one with higher significance is chosen. To give an example, in Fig 6-1, we detected four clusters due to the existence of 5 firms, as shown in (a). Instead of including all clusters into our statistical models, we only keep the one with the highest significance (the dark circle as shown in (b).

The two firm clusters chosen for the in analysis in this chapter: urbanization clusters and localization cluster, within 1.75km radius range are shown as Fig. 6-2.

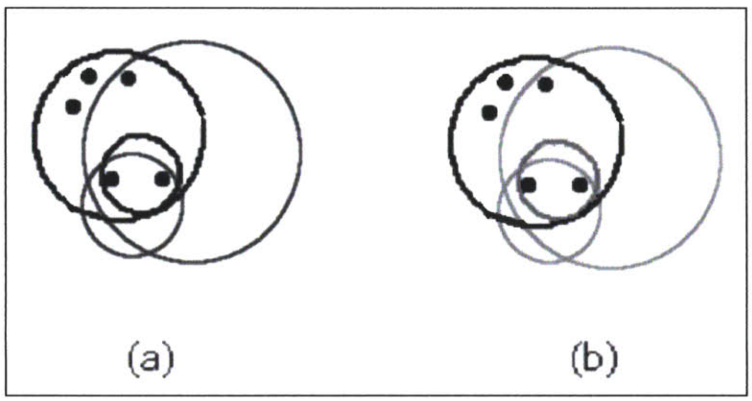
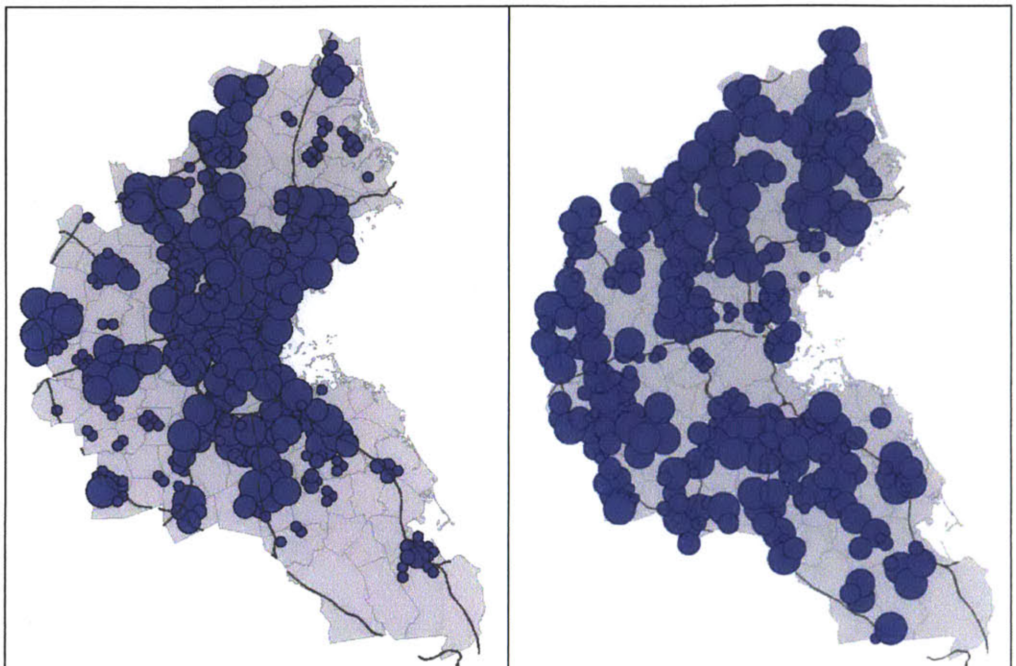


Figure 6-1 Overlapping Clusters



Urbanization Clusters at Municipal Level

Localization Clusters at Municipal Level

Figure 6-2 Firm Clusters Within 1.75km Radius Range

6.4 Data and variables

6.4.1 Data

This chapter applies the firm establishment data from ReferenceUSA database 2009 to analyze firms' clustering behavior. In total, 7828 manufacturing firm establishments, which represent about 95% of the manufacturing firm population in the database, are used for the calculation. Firm clusters come from the calculation results in Chapter 4 and Chapter 5 and with redundant ones removed. Due the size of the study area, only clusters smaller than 1.75km in radius, which equals to the average township size in the study area, are included in this chapter's modeling analysis.

6.4.2 Variables

From the dataset, we create nine variables from firms' independent characteristics to help explain firms' clustering behaviors according to the literature findings: size, age, productivity, credit rating, technical investment, rent/lease cost, utility cost, advertising cost and a dummy variable showing whether a firm is a branch office. Firm size is measured as a firm's locational employment volume in 2009. As the actual establishment date is missing for most firms in the dataset, we use years in the database as a proxy variable for firm age, which ranges from 1 to 27. The productivity of a firm is measured as the sales per employee. We use a firm's credit rating as an index of the firm's financial performance, which is acquired from the dataset. Technical investment is measured as a firm's technical spending in year 2009. A firm's rent/lease cost refers to the cost associated with the firms' usage of building and land. Utility cost and advertising cost are the actual spending a firm has on utilities or advertising, respectively.

In addition to the above variables, we create a dummy variable to indicate whether a firm has moved between 2004 and 2009. A firm is considered to have moved if it exists in the database in both years of 2004 and 2009, and its geographic location changed over 100 meters. As geocoding errors exist in the dataset, the mismatch of firm addresses does not necessarily mean that a firm has moved. Considering most geocoding errors in the data are within 100 meters range, we use 100 meters to filter the geocoding errors. In total, we identified 1558 firms that have moved during 2004-2009.

Table 6-1 introduces the variables adopted in this chapter's models.

Table 6-1 Descriptions of Explanatory Variables

Explanatory variables	Descriptions
LogSize	The logarithm of establishment employee number in 2009
LogAge	The logarithm of years the firm has been in database till the end of 2009
LogProductivity	The logarithm of firm productivity, which is result of dividing firm sales by firm employment size, both in 2009
LogCred	The logarithm of firm's credit rating in 2009
LogTech	The logarithm of firm's technical expenditure in 2009
LogRent	The logarithm of firm's rent or lease expenditure in 2009
LogUtility	The logarithm of firm's utility expenditure in 2009
LogAdvert	The logarithm of firm's utility expenditure in 2009
Branch	A dummy variable with value of 1 when firm is a branch office, 0 otherwises
Moved	A dummy variable with value of 1 if firm has moved between 2004 and 2009; 0 otherwise

6.5 Results and Discussion

6.5.1 Firm clustering determinants

We first look into how firms' characteristics associate with their locations relative to identified clusters at municipal level. Noticing that the existence of many micro-firms (with less than 10 employees) might overweight other firms' impacts in the model, for each type of clusters, we estimate the same model on all firms as well as on firms with more than 10 employees. To test how technical expenditures and productivity level jointly affect firms' locations relative to clusters, we also estimate the model with an interaction term of them as these two production factors are usually closely related. Thus, for each of the two types of clusters: localization cluster and urbanization cluster, we have four models tested: model (1) is estimated with data of all firms without interaction term; model (2) is estimated with all firms' data with the interaction term; model (3) is estimated with the data of large firms (> 10 employees) without interaction term included; model (4) is estimated with the data of large firms (>10 employees) with the interaction term considered. Results of the empirical estimation are presented in Table 6-2 and 6-3. Specifically, Table 6-2 presents the found associations between a manufacturing firm's characteristics and its location relative to detected urbanization clusters at municipal level (<=1.75km in radius). According to equation (6-4) in this chapter, variables in all of the models have been transformed into the natural logs forms, except for the dummy variables.

The Hosmer-Lemeshow test with a chi-square distribution is used as a measure of goodness-of-fit and is found highly significant at the 0.001 level in all of the models in Table 6-2 and 6-3. This suggests that the models have a good fit of data. In all models, signs of the explanatory variables are mostly as expected. in model (1) and (2) of Table 6-2, firm size, advertising expense, rent/release expense are all with statistically significant and positive coefficients, which suggests that firms found within urbanization clusters are more likely to have larger employees, pay more for advertising and rental. As the urbanization clusters are usually found at urban centers, where there exist the most intense land use biddings, firms usually pay higher cost for space usage to locate there. In Chapter 2's literature review, we introduced that regional level studies argue that big firms tend to not cluster since they internalize most of external economies. This differs from what we find here: the big firms are more likely to be found within urbanization clusters. Although small firms might be very interested in locating within firm clusters, the higher rental cost there might keep them from doing that as they have less bidding capacity as big firms do. In Table 6-2, branch dummy variable is also with significant positive coefficient for all firms, but with insignificant coefficients for large firms (>10 employees), suggesting that small branches tend to concentrate in urbanization clusters, but branches with bigger scale (>10 employees) do not show this tendency. The productivity, utility, firm credit and firm age are not statistically different from zero in the models, although in model (1) and (3) of Table 6-2 productivity has a positive sign, which suggests a positive association between firm productivity and locating within urbanization clusters, but this association is not significant and unstable.

Table 6-3 presents the results of empirical estimation about the association between firms' characteristics and their locations relative to localization clusters at municipal level (<1.75km in radius). In model (5) and (7), firm size, technical expense and firm age are all with significant positive coefficients, suggesting that firms found within localization clusters are with larger employment size, higher technical expense and longer operating years. Productivity is significant and positive in model (5) for all firms but becomes not statistically different from zero in model (7), which indicates that for small firms, locating within localization clusters is associated with positive productivity improvement, but for large firms (>10 employees), locating within localization clusters does not cause significant productivity growth. From both models of (5) and (7) in Table 6-3, we observe significant and negative coefficients for utility. That indicates the firms within localization clusters bears less utility cost compared with other firms. In model (5), coefficients of rental cost and advertising cost are significant and negative, but become not statistically different from zero in model (7). This suggests that while small firms are found with less rental and advertising expense, it is not true for large firms (>10 employees). The dummy variable branch is not significant in models of all firms, but is significant and negative in models of large firms, suggesting that large branch firms are less likely to be found within localization clusters.

Table 6-3 shows that locating within localization clusters brings firms various benefits like savings from utility and advertising, higher productivity, lower rental cost, and larger employment. For a firm, a one-percent increase in firm size, or increase in log size of 0.01, increases log odds of locating within localization clusters by 0.00742 (it is the product of firm size's coefficient and 0.01). For the sample data, 47.8% firms locate within clusters (the odds of firms locating within clusters is 47.8/52.2=0.917), so a one-percent increase in firm size increases the odds of a firm locating within clusters to 48.15% (it is the result of 0.478*e0.00742). Using the same calculation, we can deduce the quantitative extent of other coefficients. Table 6-3 also indicates active technical expenditure within localization clusters, which eventually improves firms' technical level and helps promote productivity. The positive and highly significant coefficient of technology in the firm clusters suggest that firms are locating within clusters characterized by a high level of technological consumption. These findings suggest the economic value of clusters as an access to knowledge resources that in turn manifests itself in higher firm productivity within the clusters. This is in line with theoretical models such as Ericson and Pakes (1995), Dosi et al. (1995) and Ballot and Taymaz(1997) .

Comparing the observations from Table 6-2 with those from Table 6-3, we find that large firms are more likely to be found within both types of clusters and firms are found to pay more for rental in urbanization clusters, and pay less in localization clusters, compared with those firms that are outside of clusters. Firms within localization clusters are found with less utility cost and higher technical expense, but this does hold for firms within urbanization clusters. While older firms are more likely to be found within localization clusters, firm age is not significant in models of urbanization clusters.

To compare the magnitude of explanatory variables' association with firms' location relative to clusters, Chart 6-1 gives the comparison of standardized coefficients of model (1), (3), (5),(7). The standardized coefficients is the estimated coefficients based on standardized variables, which is with a mean of 0 and standard diviation of 1. As the standardized coefficients for interaction terms are not interpretable, those models with interaction terms are not graphed here. From the chart, we can see that for both types of clusters, rent and firm size all significantly and positively associate with firms' location within clusters and rent is actually the biggest explaining factor of firms' location differences relative to clusters. In models of localization clusters, utility and advertise are among the major explaining factors too but, contrary to models of urbanization clusters, they negatively associate with the dependent variable. From models (3), (5) and (7), we can see that firms in clusters are with higher productivity larger technical expense.

Chart 6-1 Relative Importance of Variables

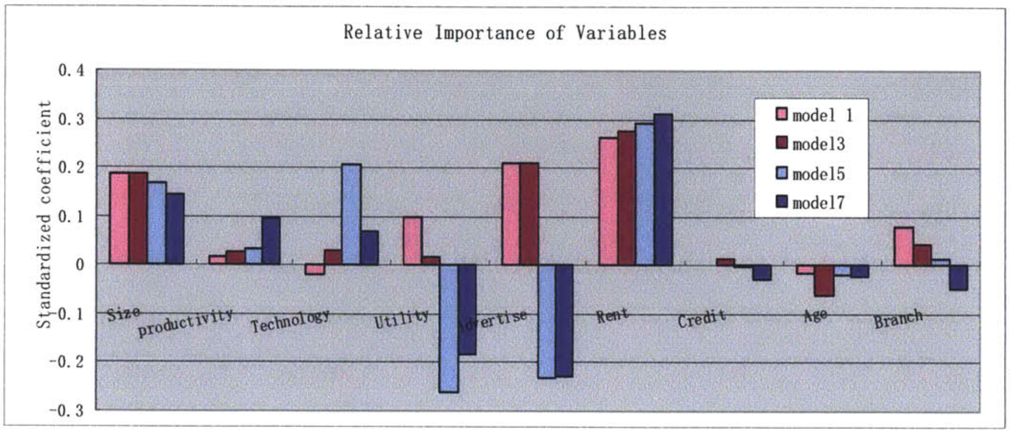


Table 6-2 Results of Urbanization Clusters

Explanatory Variables	Models with all firms Estimates (standard error)		Model of firms with more than 10 employee	
	Model 1	Model 2	Model 3	Model 4
Intercept	0.414(0.719)	1.884(0.947)	0.0419(1.4191)	2.279(1.839)
Log_Size	0.612(0.066) ***	0.625(0.067) ***	0.852(0.202) ***	0.855(0.202) ***
Log_productivity	0.048(0.056)	-0.068(0.074)	0.074(0.103)	-0.101(0.137)
Log_Technology	-00019(0.118)	-0.961(0.421) *	0.028(0.257)	-1.252(0.723) *
Log_Utility	0.093(0.106)	0.142(0.107)	0.015(0.235)	0.087(0.236)
Log_Advertise	0.202(0.108) *	0.236(0.109) *	0.193(0.170)	0.229(0.171)
Log_Rent	0.238(0.100) *	0.268(0.101) **	0.235(0.156) *	0.270(0.157) *
Log_Credit	0.010(0.086)	0.0006 (0.086)	0.059(0.166)	0.058(0.166)
Branch	0.431(0.114) ***	0.468(0.116) ***	0.202(0.189)	0.266(0.192)
Log_Age	-0.035(0.041)	-0.040 (0.041)	-0.153(0.098)	-0.172 (0.098) *
Productivity* Technology		0.070(0.030) *		0.009(0.050) *
R^2 (Cox & Snell)	0.1246	0.1254	0.1178	0.1196
R^2 (Nagelkerke)	0.1389	0.1403	0.1314	0.1347
χ^2 (df)	540.9137(9)***	544.863(10)***	330.815(9)**	333.8193(10)**

Note: Independent variable: logit of firm cluster occurring

*** $\rho<0.001$

** $\rho<0.01$

* $\rho<0.1$

Table 6-3 Results of Localization Clusters

Explanatory Variables	Models with all firms Estimates (standard error)		Model of firms with more than 10 employee	
	Model 5	Model 6	Model 7	Model 8
Intercept	-2.743 (0.568) ***	-0.4869 (0.778)	-1.9138 (1.0003)*	1.4267 (1.3309)
Log_Size	0.742 (0.051) ***	0.7601 (0.0515)***	0.3282 (0.1242)**	0.3313 (0.1245)**
Log_productivity	0.137 (0.044) **	-0.0421 (0.0609)	0.1197 (0.0730)	-0.1418 (0.1003)
Log_Technology	0.445 (0.093) ***	-0.8644 (0.3238)**	0.5006 (0.1815)**	-1.3039 (0.5078)*
Log_Utility	-0.491 (0.084) ***	-0.4267 (0.0855)***	-0.5547 (0.1669)***	-0.4687 (0.1675)**
Log_Advertise	-0.254 (0.088) ***	-0.2029 (0.0889)*	-0.0535 (0.1318)	0.00656 (0.1323)
Log_Rent	-0.267 (0.082) **	-0.2183 (0.0827)**	-0.0725 (0.1220)	-0.0162 (0.1225)
Log_Credit	0.038 (0.067)	0.0283 (0.0673)	0.0348 (0.1128)	0.0384 (0.1125)
Branch	-0.019 (0.080)	0.0281 (0.0814)	-0.2896 (0.1229)*	-0.2205 (0.1248)*
Log_Age	0.085 (0.034) *	0.0760 (0.0336)*	0.1240 (0.0672)*	0.0963 (0.0677)
Productivity* Technology		0.0970 (0.0230)***		0.1341 (0.0353)***
R^2 (Cox & Snell)	0.1597	0.1623	0.1239	0.1314
R^2 (Nagelkerke)	0.1798	0.1833	0.1323	0.1423
χ^2 (df)	383.5668 (9)	401.5905 (10)	146.4303(9)	161.1856(10)
P	<0.0001	<0.0001	<0.0001	<0.0001

Note: Independent variable: logit of firm cluster occurring
R^2= 0.1597 (Cox & Snell), 0.1798 (Nagelkerke), ρ<0.0001.
*** ρ<0.001, ** ρ<0.01, * ρ<0.1

6.4.2 Does productivity and technology jointly affect firm location?

Technology is a crucial factor affecting productivity. In fact, it is the only source of permanent increases in productivity. We have good reasons to believe that technology and productivity is affecting firms' location behaviors jointly. We have therefore extended the models to include an interaction term of productivity and technology. The results are shown in models (2), (4), (6) and (7) for urbanization clusters and localization clusters respectively. The models for both types of clusters show statistically significantly coefficients for the interaction term of productivity and technical expense, suggesting that a firm's location relative to clusters is significantly associated with the combined impacts of its productivity level and technical expense.

As can be seen from Table 6-2 and 6-3, models (2) and (4) with interaction term included perform better than models (1) and (3), which do not include the interaction effects. The Wald statistic of the interaction term, which is a Z, is calculated by dividing coefficient by its standard error (showing in the parentheses) and it is highly significant at the 0.001 level. Compared with models without interaction term, we observed larger R-square values in the models with interaction terms. That indicates that inclusion of the interaction term increases the explaining capacity of our model.

The coefficient in the models with interaction terms can be interpreted as the amount of change in the slope of the regression of logit on productivity when technical expense changes by one unit. In integrating the estimated coefficients for them, we need to consider both major effects and joint effects. Using Table 6-3 model (6) as example, assuming one unit increase of log_productivity and one unit increase of technology, we can calculate the log odds of firms locating within clusters as: -0.0421+(-0.8644)+0.0970=-0.8095; if assuming one unit increase of log_productivity and no increase of technology, the log odds of firms locating within clusters change to -0.0421. Both results diverge from our discovery in models (1) and (3). This indicates a complicated interaction between the two variables: technical expense and productivity, in associating with firm clustering. To infer the degree and direction of their association with firm clustering, more detailed understanding of firms' technical expense and productivity level will be needed.

6.4.3 Why do firms leave/enter clusters?

Firms relocate for various reasons – to accommodate growth, to save money, to be closer to vendors and clients, or just to enjoy a nicer natural environment. When firms decide to relocate, they make their decisions based on good reasons that optimize their benefits and meet their interests. Therefore, the location patterns of those relocated firms can imply extra messages about firm clustering. Using the data of firms that relocated between 2004 and 2009, we look into the relationship between the probability of a firm's leaving or entering firm clusters and its characteristics. Table 6-4 shows the empirical results of four scenarios: enter an urbanization cluster, leave an urbanization cluster, enter a localization cluster and leave a localization cluster.

In the study area, between 2004 and 2009, firms that moved into urbanization firms are found with lower rental expense, higher advertising expense and younger age; firms that moved into localization clusters are also with younger age, but with higher productivity. These results are consistent with Phelp, Fallon and Williams (2001)'s finding about new firms' favoring spatial proximity to firm clusters. Also the localization model result confirms previous model (5)'s finding that firms within localization clusters are with productivity advantages from another perspective.

Table 6-4 also shows that smaller firms are more likely to move out of firm clusters. While this result does not verify Wolf and Pett (2000)'s theory that small firms favor spatial proximity to clusters for externalities reasons, it is consistent with our findings in 6.4.2 that larger firms are more likely to be found within clusters. Additionally, from this result, we know that there exist certain mechanisms to encourage large firms moving into firm clusters, even if they did not choose to locate within firm clusters at first. Firms that moved out of localization clusters between 2004 and 2009 are also found with higher advertising expense, lower rental expense and younger age.

Table 6-4 Results of Urbanization Clusters

Explanatory Variables	Urbanization Cluster Models Estimates (standard error)		Localization Cluster Models Estimates (standard error)	
	Enter Cluster	Leave Cluster	Enter Cluster	Leave Cluster
Intercept	1.206(2.356)	-2.513(2.317)	5.097(2.707)*	3.891(2.526)
Log_Size	0.293(0.209)	-0.644(0.201)**	0.378(0.234)	-0.516(0.200)**
Log_roductivity	-0.022(0.185)	0.151(0.180)	-0.349(0.214)	-0.308(0.198)
Log_Technology	0.522(0.356)	0.175(0.393)	0.699(0.328)*	0.168(0.384)
Log_Utility	-0.3776(0.313)	-0.143(0.364)	-0.347(0.295)	-0.075(0.351)
Log_Advertise	0.955(0.364)**	-0.275(0.361)	-0. 357 (0. 388)	0.682(0.371)*
Log_Rent	-0.891(0.337)**	0.197(0.333)	0.389(0.360)	-0.723(0.347)*
Log_Credit	0.276(0.243)	-0.0246(0.275)	-0.288(0.297)	0.063(0.276)
Branch	0.456(0.434)	-0.176(0.147)	0.071(0.423)	0.430(0.305)
Log_Age	-0.499(0.164)**	-0.047(0.342)	-0.838(0.169)***	-0.555(0.138)***
R^2 (Cox & Snell)	0.0708	0.0187	0.1106	0.0456
R^2 (Nagelkerke)	0.0963	0.0339	0.1629	0.0833
χ^2 (df)	30.391 (9)***	15.641 (9) *	48.4451(9)**	39.9007(9)***

Note: *** $\rho<0.001$
** $\rho<0.01$
* $\rho<0.1$

6.6 Conclusion

The empirical data on which this chapter is based strongly suggest that firms within clusters are significantly different from those outside of clusters. Firms with larger employment size, productivity and rent expense are more likely to be found in either urbanization clusters or localization clusters. Attributes like productivity, technical expense, advertise expense, utility expense and branch or not also differently associate with firms' location relative to clusters, but the exact magnitudes and direction depends on which types of clusters is discussed. For firms in localization clusters, higher technical expense helps promote firms' technical level, which ultimately leads to higher productivity, as indicated by the results. Firms within localization clusters are also found paying less utility expense, which can be the result of infrastructure sharing among many firms in the same industry. In urbanization clusters, firms are found with higher utility cost and insignificant productivity advantages compared with firms outside of clusters. Branch firms are more likely to be found within clusters, except large branch firms, which are found negatively associated with localization clusters.

In models we observe significant joint association of firms' technical expense and level with their locations relative to clusters, suggesting the existence of agglomeration economies within the firm clusters. In localization clusters, the significant and positive association among technical expense, productivity level and clustering signifies the existence of localization economies, where the sharing of labor, techniques and intermediate products among firms within the clusters ultimately promote firms' productivity. But as we do not have directly proxy variables for these inter-firm interactions, our findings here are more like a general support, instead of empirical evidence.

The last part of this chapter studies the association of firms' characteristics with their location relative to clusters with the data of those relocated firms between 2004 and 2009. The results indicate that firm clustering is a dynamic process, instead of a static snapshot. Firms adjust their clustering behaviors according to their new demand. For example, we found those relocated firms with larger employee size are more likely to be found within clusters, but the smaller firms are more likely to be found outside clusters.

CHAPTER 7: SPATIAL DETERMINANTS OF FIRM CLUSTERS

7.1. Introduction

In chapter 4 and 5, we introduced the methods of detecting two types of firm clusters: cluster of localization and cluster of urbanization, presented the distribution of firm clusters within the Boston metropolitan area and showed that the occurrence of firm clustering within the study area is far from randomness: firms tend to concentrate in certain areas and different types of firms exhibit divergent clustering patterns. In this chapter, we are going to take a closer look at the locational characteristics of detected firm clusters.

While firms make decisions about where to locate, various factors are considered in the process: clients, inputs, cost, safety, availability of space, traffic, etc. Among these considerations, many are location bounded, like taxes, rent, distance to highways, labor market, etc. It is obvious that not every location can become a firm cluster, because not all places provide same accessibility to these resources. Thus, in this chapter, by examining the distribution of firm clusters, we hope to provide insights into the understanding of firms' clustering behaviors and offer policy references for urban planners in related work.

7.2. Literature review

Researchers have studied industrial clustering for more than a century. Marshall (1890) provided the earliest economic analysis of industrial clustering, arguing that firms in the clusters gain productivity advantages from three micro-mechanisms: scale economies of intermediate goods, labor pooling, and knowledge spillover. Ohlin (1933) suggested dividing the agglomeration economy, which explains the industrial clustering, into three categories: economies of scales within the firm, localization economies, and urbanization economies. Since then, extensive empirical efforts have appeared in formalizing these agglomeration economies and estimating their impacts on industrial clustering, including Sveikauskas (1975), Moomaw (1981), Henderson (1986), Porter (1990), Glaeer et al. (1992), Henderson, Kuncroro, and Turner (1995). More recently, Rosenthal and Strange (2003) discovered that agglomeration economies attenuate rapidly the first few miles. Their findings question previous studies that have been forced to assume that agglomeration economies are club goods that operate at a metropolitan scale, and suggest that agglomeration should ideally be studied at a much more refined geographic level than has been the norm.

An unanswered question in Rosenthal and Strange's study is that if agglomeration exists at a small geographic level and the metropolitan area should be viewed as a whole, where will industrial clusters choose to be? The discontinuous distribution of industrial clusters can be seen in cities in forms of industrial zones, and industrial gardens or street corners, which obviously have different land use types and landscapes from other portions of the cities. There can be two answers for this question. One is that every location within the metropolitan area can become an industrial concentration spot, but that the random hand of history chose existing industrial clusters as the concentration spots. The other answer is that locations do not have identical characteristics and some offer a more productive environment. In their study, Rosenthal and Strange referenced Saxenian's point (1994): "a local industrial system has three dimensions: local institutions and culture, industrial structure and corporate organization." However, Saxenian's theory does not explain whether or how geography relates with industrial clusters.

Therefore there is a gap in literature about the impacts of geographic characteristics on agglomeration, which becomes the aim of this chapter: by examining the existence of industrial clusters and local geographical characteristics that associate with firm location considerations, this study estimates the determinants of a particular spot being occupied by industrial clusters at certain spatial levels as functions of the local geographic characteristics.

7.3 Methodological Issues

7.3.1 Selection of Firm Clusters for the Study

As discussed in previous chapters, the whole metropolitan area is covered by a hierarchical system of geographically overlapped clusters. Whichever spatial scales of clusters we choose to focus on, we are going to miss those that are unselected. This will affect the comprehensiveness of our conclusions. But if we expand the studied spatial scales, we are going to lose data variations. This will hurt the representation of the results. For example, if we choose to examine firm clusters within the city of Cambridge, we are going to miss those clusters that are bigger than Cambridge. But if we choose the Boston metropolitan area as the spatial scale, we are going to have only one sample, which is the whole study area that covers all firms and all area. In the end, the spatial scales of firm clusters have to relate with the spatial scales of study areas. Considering the focused contribution of this study is the intrametropolitan level, in this study we are going to study the firm clusters of 1750 meters in radius or less, which is the average city size in the study area.

By now, all the firm clusters are defined as the concentration of firm establishments. A one-employee firm is recognized as the same as a firm with thousands of employees. But in reality, firm size often plays an important role in firms' impacts on local economics and their clustering behaviors. To adjust for firm size, we import a new type of cluster: **firm cluster defined by employment**, which refers to the significant concentration of firm employees. For description convenience, we refer to previous clusters as **firm cluster defined by establishment**, which refers to a significant group of firm establishments. The detection of firm clusters defined by employment is basically the same as we described in chapter 4 and 5 about how to detect firm clusters defined by establishment with treating each employee as a one-employee firm establishment at the same location. While the cluster of localization defined by establishment refers to the significant firm clusters controlling for the distribution of all firm establishments, the cluster of localization defined by employment refers to the significant firm clusters controlling for the distribution of all employees in the study area.

Thus, we have four firm clusters constructed for our analysis in this chapter: cluster of urbanization defined by establishment, cluster of urbanization defined by employment, cluster of localization defined by establishment and cluster of localization defined by employment. The distributions of these four firm clusters within 1.75km radius range in the study area are shown as Fig 7-1.

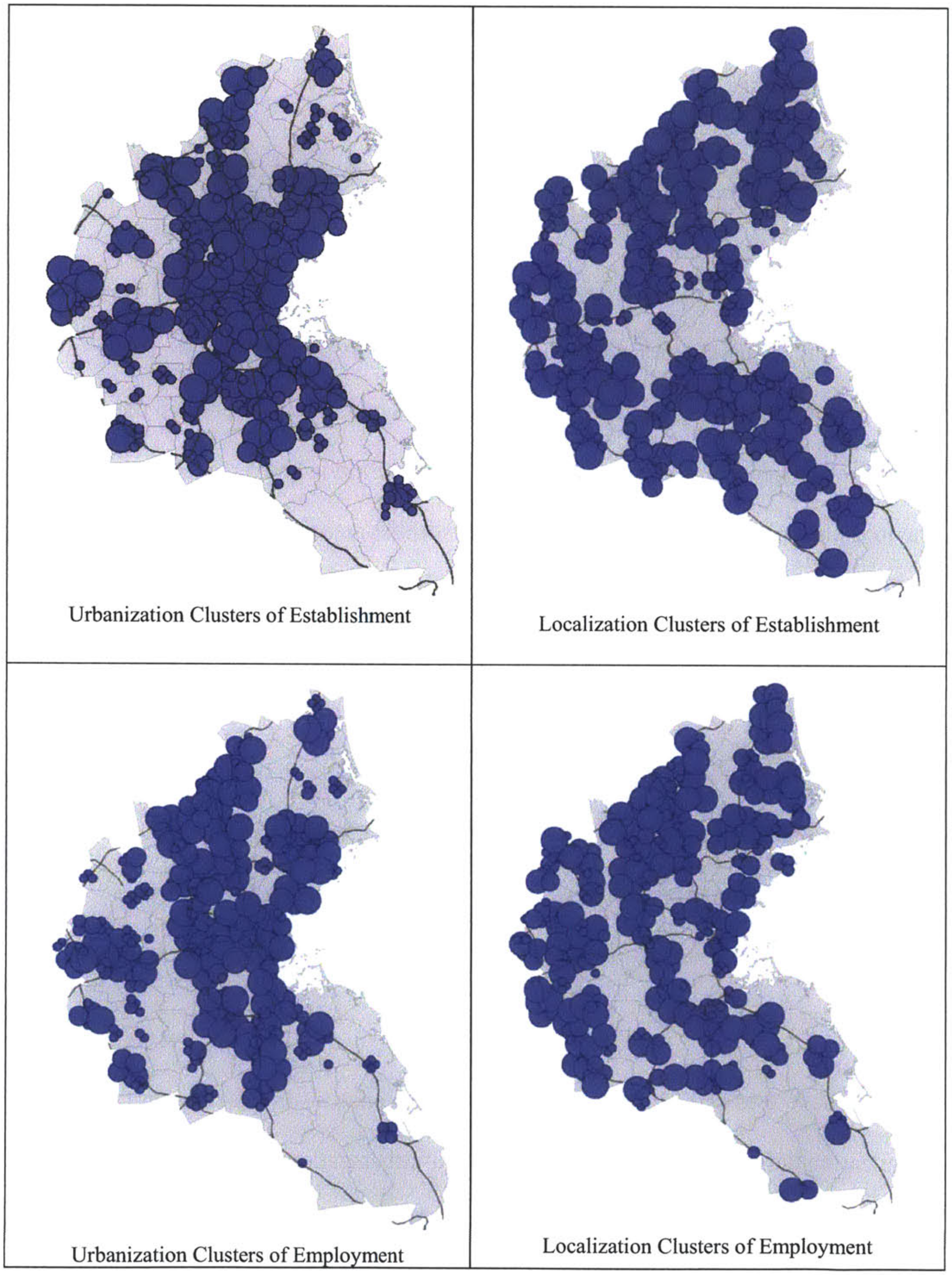
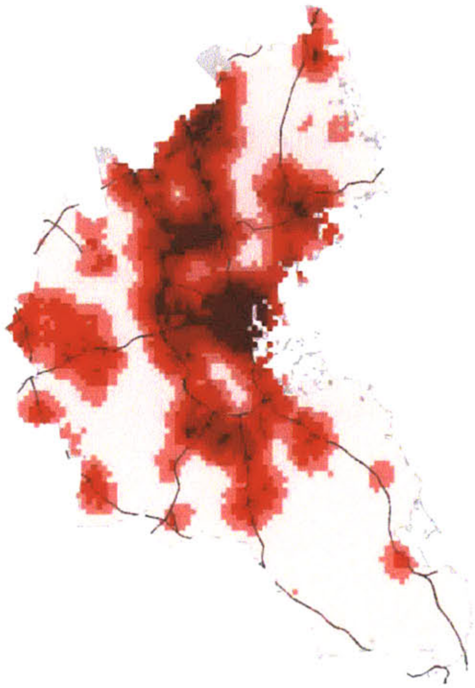
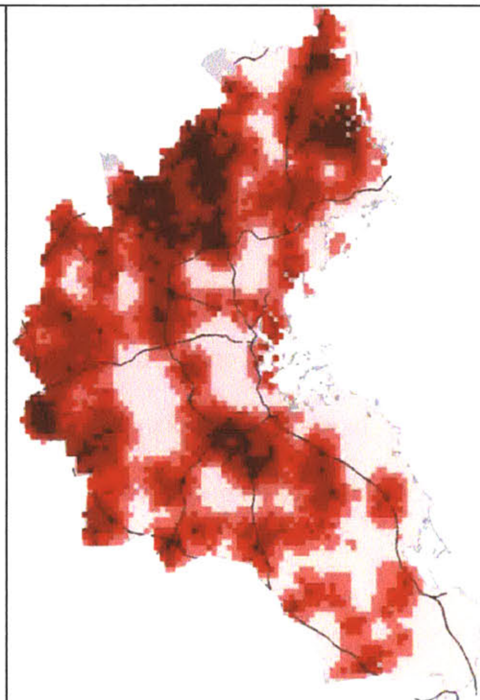


Figure 7-1 Firm Clusters within 1.75km Radius Range

We noticed that by imposing an overlapping restriction on clusters, we will filter out some firm clusters, which contain over 50% overlapping area with other more significant clusters. To adjust for this problem, we also look into firm cluster intensity, which is the number of firm clusters a place can observe without overlapping restriction in place. In Fig 7-2, we present the four intensity maps of firm clusters: intensity of urbanization cluster of establishment, intensity of urbanization cluster of employment, intensity of localization cluster of establishment, intensity of localization cluster of employment.

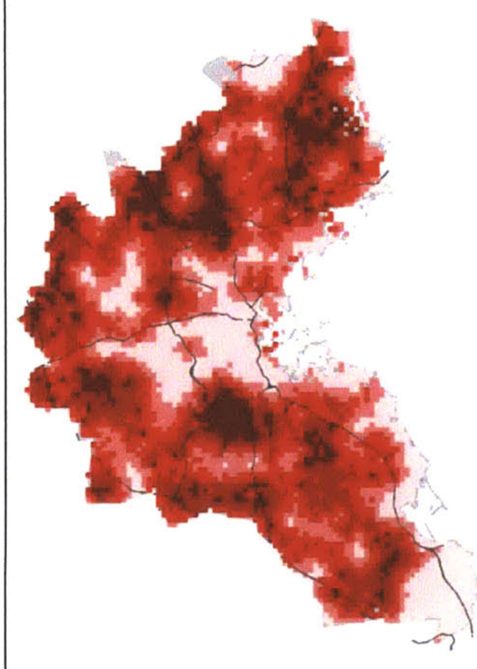


Intensity of Urbanization Cluster of Establishment



Intensity of Localization Cluster of Establishment

Intensity of Urbanization Cluster of Employment



Intensity of Localization Cluster of Employment

Figure 7-2 Intensity of Firm Clusters within 1.75km Radius Range

7.3.2 The Econometric Model

A cluster's location is an outcome of the group decision of the firms within the cluster. Firms are attracted by a site's locational convenience and externalities and bid to locate within it. Some firms get in and some don't. Our hypothesis here is that the attributes of a location represent the most important factors that explain whether a location becomes a firm cluster or not. Locations with good transportation accessibility, larger labor pool, bigger market access, cheaper land cost and lower tax burden have a better chance to attract firms and become clusters, and vice versa. As discussed in Chapter 6, in this circumstance, traditional ordinary least square (OLS) regression is not appropriate anymore since the dependent variable, whether a location is within firm clusters, can only take one of two exclusive values: one or zero, which is not continuous variable and cannot be negative. We are going to apply logistic regression to model this event and the model can take the following form:

$$P(y_i = 1) = \frac{1}{1 + \exp^{-f(X)}},$$

$$\text{or } \log it(P) = \frac{P(y_i = 1)}{1 - P(y_i = 1)} = f(x), \quad (7\text{-}1)$$

where $P(y_i = 1)$ is the probability that a location is chosen as a firm cluster, **X** is a location's attributes that is specific to a location. In literature, **X** is estimated with different forms. In this study, we assume $f(x)$ can be defined as follows:

$$f(x) = K_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_m X_m \; , (7\text{-}2)$$

where K, $\alpha_1 \ldots \alpha_m$ are unknown constants and $X_1 \ldots Xm$ are locational attributes for a site.

Combining (1) and (2), we have the general model that is going to be used in this chapter:

$$\log it(P) = \ln K_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_m X_m \qquad (7\text{-}3)$$

As the intensity of firm clusters is measured as the number of overlapping firm clusters that can be observed as a location, it is a continuous positive number. We can use traditional OLS regressions to model its relationships with independent variables.

$$I = K_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_m X_m \qquad (7\text{-}4),$$

where I is the intensity vale of a location and $X_1 \ldots Xm$ are defined as equation (7-2).

7.3.3. The Autologistic Autocorrelation

Another methodological difficulty comes from the spatial autocorrelation in the spatial distribution of firm clusters. On one hand, spatial autocorrelation may arise from the "intrinsic" spatial structure (Anselin 1992) of firm clusters because the potential locations for industries are limited and sites around those areas will have a better chance to become a firm cluster, and vice versa. On the other hand, spatial autocorrelation may arise from environmental "forcing" (Legendre 1993), whereby the determinants of clusters are spatially structured, such as industrial land, roads, natural amenity, etc. The locations of firm clusters are essentially categorical: presence/absence. In this study, firm clusters are easily fitted using logistic regression to study their locational characteristics. Most methods for modeling spatial autocorrelated data are restricted to continuous rather than categorically-dependent data. This derives from the mathematical intractability of likelihood functions for discrete-variable auto-models, in contrast to that of continuous-variable, for which a maximum likelihood approach is relatively straightforward (Besag 1974). We address this problem by adding an autocovariate term to the logistical regression model, resulting in the autologistic model (Besag 1972;1974).

A logistic regression model like equation (7-3) is straightforward to fit using maximum likelihood methods (Collett, 1999) with the notorious assumption of independence of irrelevant alternatives (IIA). The IIA property ensures a consistent estimate of a logistic model from a reduced choice set. When IIA is violated, reduction of alternatives would lead to a loss of valuable information, thus biased estimation. In our study, IIA becomes a problem because the geographic attributes of firm clusters are essentially correlated because of the continuous distribution of geography and they are often the influential criteria of location decisions. Such elements as tax, highway access, and land availability are often location dependent and are often the important consideration in firm's site selection. Moreover, we use discrete space (1k*1k grid with unit of meter) to characterize the set of spatial alternatives available for industrial clusters and the universe of choice is assumed to be cellular locations within the metropolitan area. The spatial dependency among alternatives becomes more obvious as one site with similar geographic characteristics usually contains multiple choice sets. We solve the auto-correlation problem by using the autologistic model, which contains an added autocorrelation, here called autocov, and leads to the model (Besag 1972;1974):

$$\log it(P_i) = \alpha + \beta * X_i + \gamma * auto\,\text{cov}_i + \varepsilon_i \qquad (7\text{-}5),$$

where

$$auto\,\text{cov}_i = Wi * \log it(P_i),$$

W_i is a weight matrix, defining the spatial relation between a site and its neighbors. Commonly there are four arrangements used to define the neighbors included in the considerations, referred to as nearest neighbors (first-order neighborhood) or next-to-nearest neighbors (second-order neighborhood), and the like (fit) (Wintle and Bardos, 2006). As shown in Fig 7-3, the white cells represent those included in the neighborhood of the central cell.

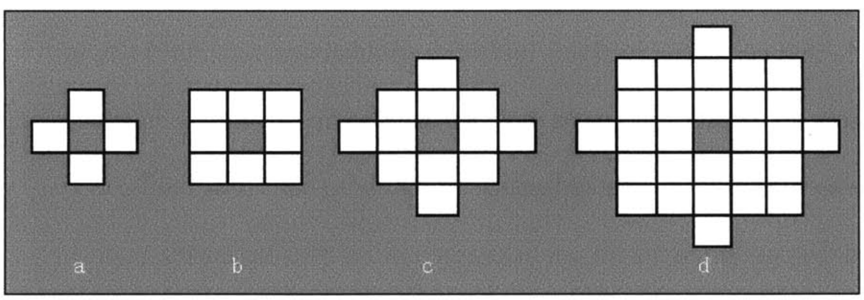


Figure 7-3 (a) first-order, (b)second-order, (c)third-order, and (d) forth-order neighborhoods (Duplicated from Wintle and Bardos, 2006)

In this study, since the exact range of spatial autocorrelation is unclear, we construct four scenarios of neighbors to test the impacts of spatial autocorrelation: first-order, third-order, fifth-order and seventh-order.

7.4. Data and Variables

7.4.1. Data

Three groups of datasets were drawn for the analysis in this chapter. Detected industrial clusters by using firm establishment datasets from ReferenceUSA database were used to calculate the dependent variable Pi; spatial datasets from MassGIS were applied to calculate the locational attributes of each grid within the Boston metropolitan area, including highway, zoning, land use, roads, towns, universities, coastal lines; another dataset from MassGIS, vehicle miles traveled (vmt) datasets from MassGIS was used to provide work trip, population and employment information of each locations.

The study region, Boston metropolitan area, is converted into 1000m*1000m grids. Other data layers are superimposed upon the grid layer and the shortest straight line distance from each grid's center to all other spatial considerations is calculated as independent variables.

7.4.2 Variables

For each grid cell, we calculate four groups of locational variables to evaluate the impacts of geography on the appearance of industrial clusters: accessibility, land availability, labor pool and business environment. For accessibility, we measure the straight line distances from each grid cell to the closest spatial objects including highway exits, highway intersection downtown Boston, coastal line, and major roads. For land availability, we calculate the developed land ratio within the cell. For the labor pool, we measure total population, total employment within each cell, and we also calculated the available labor within the commuting distance by using the VMT data from MassGIS. For the business environment, we use industrial tax as a proxy variable business environment. Distance to colleges was also used as a proxy for knowledge accessibility, but as most of the colleges in the Boston Metro mainly locate around central area, distance to colleges showed a high correlation (0.624) with distance to downtown. It was finally removed from the analysis to avoid multicollinearity problems within the models.

The definitions of explanatory variables and their data sources are displayed in Table 7-1 and the Table 7-2 presents descriptive statistics of explanatory variables.

Table 7-1 Definition of Explanatory Variables

Explanatory variables		Descriptions	Sources
Name	Variable		
Land development ratio	Ldr	The percentage of land is already developed	MassGIS land use layer 2005
Proximity to highway exit	HwyExit	-1 * Distance to closest interstate highway exits (unit: km)	MassGIS road layer 2009
Proximity to highway intersection	HwyInt	-1 * Distance to closest interstate highway exits (unit: km)	MassGIS road layer 2009
Proximity to downtown	Dtown	-1 * Distance to Boston Downtown (unit: km)	
Proximity to Coast	Coast	-1 * Distance to closest coast lines (unit: km)	
Proximity to Major road	Road	-1 * Distance to closest major road (unit: km)	MassGIS road layer 2009
Office rent	Rent	Average rent expenditure of firms (unit: one thousand dollars)	ReferenceUSA business database 2009
Industrial tax	Tax08	Industrial tax in 2008 (unit: percentage)	Massachusetts Division of Local Services Gateway
Employment	Empmt	Total employment demand within the cell (unit: one)	MassGIS VMT Data 2000
Labor pool	Labor	Total available labor supply within commuting range (unit: one thousand)	MassGIS VMT data 2000

Table 7-2 Explanatory Variables Descriptive Statistics and Correlation Matrix

Type	Name	Empmt	Labor	HwyExt	HwyInt	Dtown	Coast	Mjroad	Tax	Rent	Ldr
MEAN		17.300	76.508	3.440	10.472	32.962	18.945	1.507	11.007	22.413	0.422
STD		126.515	34.447	2.635	6.169	14.270	12.429	1.546	1.784	62.378	0.269
CORR	Empmt	1.000	0.083	0.099	0.088	0.086	-0.018	0.011	-0.009	0.399	0.119
CORR	Labor	0.083	1.000	0.152	0.335	0.461	0.168	0.065	-0.070	0.175	0.499
CORR	HwyExt	0.099	0.152	1.000	0.411	0.142	-0.062	0.095	-0.100	0.203	0.258
CORR	HwyInt	0.088	0.335	0.411	1.000	0.432	-0.322	-0.019	0.059	0.177	0.300
CORR	Dtown	0.086	0.461	0.142	0.432	1.000	0.060	0.066	0.003	0.182	0.520
CORR	Coast	-0.018	0.168	-0.062	-0.322	0.060	1.000	0.016	-0.530	-0.015	0.090
CORR	Mjroad	0.011	0.065	0.095	-0.019	0.066	0.016	1.000	0.030	0.056	0.177
CORR	Tax08	-0.009	-0.070	-0.100	0.059	0.003	-0.530	0.030	1.000	-0.059	-0.129
CORR	Rent	0.399	0.175	0.203	0.177	0.182	-0.015	0.056	-0.059	1.000	0.264
CORR	Ldr	0.119	0.499	0.258	0.300	0.520	0.090	0.177	-0.129	0.264	1.000

7.5 Results and Analysis

7.5.1. Spatial Determinants of Firm Clusters

Following econometric equation (7-3), four logistic models are constructed and tested to explore the associations between explanatory variables and the four dependent variable: whether a location falls in cluster of urbanization defined by manufacturing establishment (for model CUE), whether a location falls in cluster of urbanization defined by manufacturing employment (for model CUM), whether a location falls in cluster of localization defined by manufacturing establishment (for model CLE), whether a location falls in cluster of localization defined by manufacturing employment (for model CLM). Please reference Fig. 7-1 for the map of four firm clusters' spatial distribution in the study area. The four models with all 10 explanatory variables included are estimated by SAS 9.0 and the results are shown in Table 7-3.

Table 7-3 Model Estimation Results of Equation (7-3)

Explanatory variables	Model CUE Estimates (standard error)	Model CUM Estimates (standard error)	Model CLE Estimates (standard error)	Model CLM Estimates (standard error)
Intercept	1.977*** (0.399)	0.280 (0.400)	2.9586*** (0.3292)	0.824* (0.302)
Empmt	0.015***(0.003)		0.002*(0.0008)	
Labor	0.018*** (0.002)	0.029*** (0.002)	-0.008*** (0.001)	0.008*** (0.001)
HwyExit	0.322***(0.020)	0.273*** (0.019)	0.198** (0.014)	0.096*** (0.013)
HwyInt	0.070***(0.009)	0.029** (0.009)	-0.020*(0.007)	-0.030*** (0.006)
Dtown	0.038*** (0.005)	0.038*** (0.006)	0.051*** (0.004)	0.006*(0.003)
Coast	-0.029*** (0.004)	-0.029*** (0.005)	-0.061*** (0.004)	-0.057*** (0.003)
Road	0.011 (0.026)	0.118 *** (0.027)	0.069* (0.021)	0.098*** (0.019)
Ldr	1.350*** (0.179)	2.792*** (0.1838)	-0.218 (0.148)	-0.838*** (0.139)
Tax08	-0.177*** (0.025)	-0.174*** (0.0253)	-0.116** (0.021)	-0.141*** (0.020)
Rent	-0.00007 (0.001)	0.006*** (0.0008)	0.003** (0.001)	0.004*** (0.0006)
Model Fit Statistics				
R^2 (Cox & Snell)	0.3988	0.4536	0.1877	0.0995
R^2 (Nagelkerke)	0.5383	0.6072	0.2503	0.1336
Wald $\chi 2$	1240.931***	1433.884***	753.9379***	476.888***

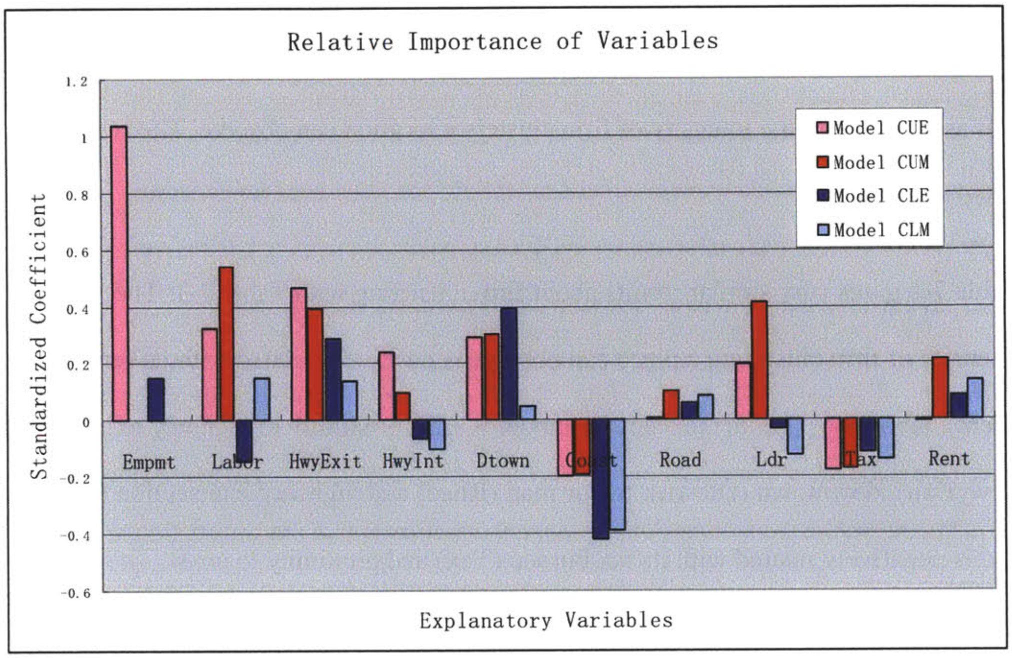
*** 0.0001
** 0.001
*0.01
CUE: cluster of urbanization defined by manufacturing establishment
CUM: cluster of urbanization defined by manufacturing employment
CLE: cluster of localization defined by manufacturing establishment
CLM: cluster of localization defined by manufacturing emloyment

The Wald test (chi-square test with the degree of 10 or 9), similar to the F-test linear regression, is used to check the validity of the model and it shows that our models are all significant at the 0.0001 level. Two pseudo R^2, Cox & Snell R^2 and Nagelkerke R^2, are used to measure how well the models fit the data. The R^2 shown in the results table indicates that models of urbanization perform much better than models of localization in explaining a location's chance of being within a firm cluster. This can be explained with the differences between the distribution of urbanization clusters and localization clusters. From Fig 7-2, we can see that clusters of urbanization mainly concentrate in central areas of the study region, but the clusters of localization distribute in a much wider areas, which obviously bring more variations into the model. This can also be confirmed by the estimates of the variable "land development ratio": those locations with higher development ratio are more likely to be found as clusters of urbanization, but less likely to be as clusters of localization.

From Table 7-3, we can see that although clusters of urbanization and clusters of localization have obviously different distribution pattern, they all show preferences in those locations with lower industrial taxes, better proximity to major road, downtown and highway exists. Firm clusters also try to avoid coastal areas, where exist strong bids for luxury residential development. Two variables are with opposite signs in models of urbanization clusters (CUE and CUM) and models of localization clusters (CLE and CLM): proximity to highway intersections (HwyInt) and land development ratio (Ldr). While urbanization clusters locate at those locations with good multiple highway accessibility and higher land development ratio, localization clusters tend to locate away from these areas. In literatures, researchers have noticed that localization economy and urbanization economy coexist to sustain the agglomeration economies and cities, but few have noticed that these two economies will cause different spatial patter of industries with urbanization firm clusters in the "core" areas and localization firm clusters appear in "periphery" areas, which in the end arises from firms' different bidding capacity and locational demand.

To compare the relative importance of the explanatory variables, we also calculated the standardized coefficients for the four models: CUE, CUM, CLE and CLM. The results are shown as Chart 7-1. From the chart, we can see that locational employment (Empmt), Labor, land development ratio (Ldr) and proximity to highway exits are the top positive factors for a location's being within urbanization clusters; but the proximity to coastal lines (Coast) and tax burdens (Tax) are the major negative factors. For a location's being in localization firm clusters, the proximity to downtown (Dtown), proximity to highway exit (HwyExit) are the top positive factors, Coast, Tax and HwyInt are the major negative factors. The labor pool (Labor) is the only directional difference between the two localization cluster models: CLE and CLM. For CLE, Labor is a major negative factor, but for CLM, the opposite holds. As CLM is defined by manufacturing employment and CLN by manufacturing establishments, this difference discloses the inconsistency between the distribution of firm establishments and firm employments: while firm employment centers naturally locate close to labor pool, firm establishment centers locate away from these locations to avoid land bid competition of residential usages.

Chart 7-1 Relative Importance of Variables in Firm Cluster Models



7.5.2. Spatial Determinants of Firm Cluster Intensity

Using equation (7-4), we constructed four models to measure the spatial determinants that associate with the intensity of firm clusters a location can observe. The results are listed as Table 7-4.

Table 7-4 gives very similar results about firm clustering with Table 7-3. The intensity of firm clustering a place can observe is positively related with its rent (Rent), labor pool (Labor) (except for model ICLE), proximity to highway exit (HwyExit), downtown (Dtown), major road (Road) and highway intersection (HwyInt) and is negatively related with its tax burden (Tax) and proximity to coast. The two urbanization models (ICUE, ICUM) and the two localization models (ICLE and ICLM) are found with opposite signs in the estimates of land development ratio (Ldr), which discloses that while clusters of urbanization are more observed at highly developed places, firm clusters of localization tend to locate at less developed areas. But as accessibility to downtown is positively significant in all four models, firm clusters just choose less developed locations that with good downtown proximity to avoid the bidding competitions of other land use purposes.

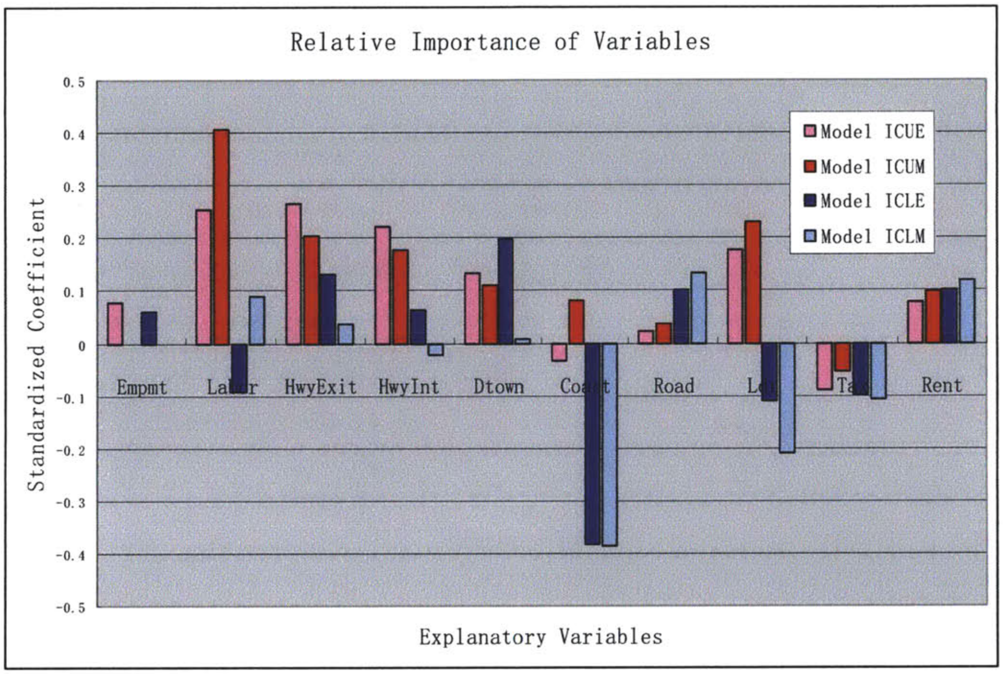
Table 7-4 Model Estimation Results of Equation (7-4)

Explanatory variables	Model ICUE Estimates (standard error)	Model CUM Estimates (standard error)	Model ICLE Estimates (standard error)	Model CLM Estimates (standard error)
Intercept	11.370*** (0.681)	8.524*** (0.651)	17.278*** (0.847)	15.485*** (0.916)
Empmt	0.004*** (0.001)		0.003*** (0.001)	
Labor	0.056*** (0.003)	0.109*** (0.003)	-0.018*** (0.004)	0.018*** (0.004)
HwyExit	0.740*** (0.028)	0.713*** (0.027)	0.331*** (0.035)	0.099* (0.038)
HwyInt	0.264*** (0.014)	0.267*** (0.014)	0.068** (0.177)	-0.026 (0.019)
Dtown	0.069*** (0.008)	0.072*** (0.007)	0.093*** (0.009)	0.004 (0.010)
Coast	-0.020** (0.007)	0.061*** (0.006)	-0.203*** (0.009)	-0.217*** (0.009)
Road	0.108*(0.044)	0.225*** (0.042)	0.430*** (0.054)	0.605*** (0.058)
Ldr	5.016*** (0.310)	7.925*** (0.294)	-2.741*** (0.385)	-5.451*** (0.412)
Tax	-0.369*** (0.045)	-0.277*** (0.043)	-0.361*** (0.056)	-0.417*** (0.060)
Rent	0.011*** (0.001)	0.014*** (0.001)	0.013*** (0.002)	0.013*** (0.001)
		Model Fit Statistics		
R^2	0.6023	0.7549	0.2341	0.1609
R^2-adjust	0.6015	0.7545	0.2326	0.1594
F value	745.43***	1822.47***	150.47***	113.42***

*** 0.0001
** 0.001
*0.01
ICUE: Intensity of urbanization cluster defined by manufacturing establishment
ICUM: Intensity of urbanization cluster defined by manufacturing employment
ICLE: Intensity of localization cluster defined by manufacturing establishment
ICLM: Intensity of localization cluster defined by manufacturing employment

Chart 7-2 presents the relative importance of explanatory variables. From the chart, we can see that while the proximity to highway exit, downtown and road are positive factors associated with intensities of all types of firm clusters, tax burden for industries is the only negative factor. The labor exhibits a opposite effects on CLE and CLM, which demonstrates the different location preferences between firm establishments and manufacturing employments. While the center of manufacturing employments are found close to labor pools, the center of manufacturing establishments are located away from them to avoid land bidding competition.

Chart 7-2 Relative Importance of Variables in Firm Cluster Intensity Models



7.5.3 Control for Spatial Autocorrelation

Following the methods introduced in equation (7-5), we recalculated model CUMANUE, the cluster of urbanization defined by manufacturing establishment, and model CLMANUE, the cluster of localization defined by manufacturing establishment with spatial autocorrelation considered. For each model, four scenarios with different neighborhood settings: 1km, 3km, 5km and 7km are tested. The results of model CUMANUE are listed in table 7-5. And Table 7-6 gives the results of model CLMANUE.

From the result tables, we can see that the autocovariate terms are consistently significant at 0.001 level. Compared with the non-spatial models listed in Table 7-2 and 7-3, all models in Table 7-5 and 7-6 are with larger R^2 measurements, which indicate that there exist strong spatial autocorrelation in the data and inclusion of spatial autocovariate term improves the models' explanting power. After controlling for spatial terms, two models provide several important consistent messages about firm clustering. For example, Tax is found to have negative impacts in both models, which agrees with our findings in Table 7-2 without controlling for spatial autocorrelation; highway exist (HwyExt) is another positive factor for a location's being in firm clusters, but highway intersection (HwyInt), which measures the accessibility to multiple highways, works in an opposite way for firm clustering; coastal accessibility (Coast) is also found to be a negative factor for having firm clustering in place. Different from our findings in Table 7-2, after controlling for spatial autocorrelation, urbanization model CUMANUE shows location of firm clusters negatively associates with labor pool (Labor), but in localization model CLMANUE, it works the opposite way. That is to say, clusters of localization prefer to places with plenty of labors, if we control for the impacts of spatial autocorrelation.

The comparison of explanatory variables' relative contribution to the model is displayed as Chart 7-3. In this chart, firm clusters tend to locate close to other firm clusters. They often choose those locations that are with large employment (Empmt), good accessibility to highway exits (HwyExt), low industrial taxes, and distance to highway intersections (HwyInt). After meeting above conditions, if they still can choose, clusters of localization tend to be close to labor pools, clusters of urbanization stay away from them. Chart 7-3 also discloses that the spatial autocorrelation steadily decreases with the increase of spatial range. It appears to affect firm clustering at the most proximity areas, and then its impacts become weaker, but still significant. With the decreasing impacts from spatial autocorrelation, the other explanatory factors' impacts start to grow. At certain large spatial range, the impacts of spatial autocorrelation will become negligible and the major factors will become as disclosed in Table 7-2 and 7-3.

Table 7-5 Autologit Results for Model CUMANU

Explanatory variables	CUMANUE-R1 Estimates (standard error)	CUMANUE-R3 Estimates (standard error)	CUMANUE-R5 Estimates (standard error)	CUMANUE-R7 Estimates (standard error)
Intercept	-5.942***(0.850)	-5.409***(0.646)	-4.5963***(0.5579)	-3.840***(0.515)
Empmt	0.009(0.006)	0.022***(0.005)	0.021***(0.004)	0.019***(0.003)
Labor	-0.004(0.004)	-0.009**(0.003)	-0.006(0.002)	-0.0005(0.002)
HwyExit	0.008(0.038)	0.062*(0.029)	0.189***(0.025)	0.280***(0.022)
HwyInt	-0.049**(0.019)	-0.105***(0.015)	-0.098***(0.013)	-0.081***(0.012)
Dtown	0.002(0.012)	0.0001(0.007)	-0.004(0.007)	-0.009(0.006)
Coast	-0.015*(0.009)	-0.023**(0.007)	-0.026***(0.005)	-0.03***(0.005)
Road	0.024(0.056)	0.058(0.042)	0.042(0.035)	0.015(0.031)
Ldr	-0.021(0.365)	-0.041(0.264)	0.368(0.220)	0.677**(0.199)
Tax	-0.026(0.048)	-0.065*(0.037)	-0.084**(0.032)	-0.123***(0.029)
Rent	0.0001(0.002)	-0.0003(0.002)	-0.00004(0.001)	0.0002(0.001)
AutoCov	11.696***(0.450)	11.098***(0.393)	9.3763***(0.3411)	7.914***(0.328)
		Model Fit		
R^2 (Cox & Snell)	0.6705	0.6061	0.5443	0.4950
R^2 (Nagelkerke)	0.6705	0.8184	0.7350	0.6684
Wald $\chi 2$	783.2079***	1134.0252***	1300.0650***	1350.7929***

Note: Independent variable: logit of firm cluster occurring
*** $\rho<0.0001$
** $\rho<0.001$
* $\rho<0.01$

Table 7-6 Autologit Results for Model CLMANU

Explanatory variables	Radius -1 Estimates (standard error)	Radius -3 Estimates (standard error)	Radius -5 Estimates (standard error)	Radius -7 Estimates (standard error)
Intercept	-6.089*** (0.832)	-6.553***(0.596)	-5.689***(0.497)	-3.842***(0.439)
Empmt	-0.001(0.001)	0.0003(0.001)	0.001(0.001)	0.002*(0.001)
Labor	0.006*(0.003)	0.010***(0.002)	0.010***(0.010)	0.005**(0.002)
HwyExit	0.032(0.034)	0.081**(0.023)	0.132***(0.019)	0.171***(0.017)
HwyInt	-0.033*(0.017)	-0.058***(0.012)	-0.066***(0.010)	-0.066***(0.009)
Dtown	0.0004(0.011)	-0.010(0.008)	-0.013*(0.007)	-0.005(0.005)
Coast	-0.003(0.008)	-0.002(0.006)	-0.010*(0.005)	-0.020***(0.004)
Road	0.023(0.052)	0.038(0.037)	0.002(0.029)	-0.015(0.026)
Ldr	-0.984**(0.334)	-0.453*(0.232)	-0.078(0.188)	0.051(0.167)
Tax	-0.055(0.047)	-0.044(0.033)	-0.065*(0.026)	-0.098***(0.235)
Rent	0.001(0.002)	0.001(0.001)	0.002*(0.001)	0.003**(0.001)
AutoCov	12.7024***(0.481)	10.97***1(0.331)	9.076***(0.277)	7.137***(0.248)
		Model Fit		
R^2 (Cox & Snell)	0.6761	0.5824	0.4729	0.3706
R^2 (Nagelkerke)	0.9016	0.7766	0.6306	0.4941
Wald $\chi 2$	716.0927***	1167.0073***	1245.7242***	1117.5894***

Note: Independent variable: logit of firm cluster occurring
*** $\rho<0.001$
** $\rho<0.01$
* $\rho<0.01$

Chart 7-3 Relative Importance of Variables in Autologit Models

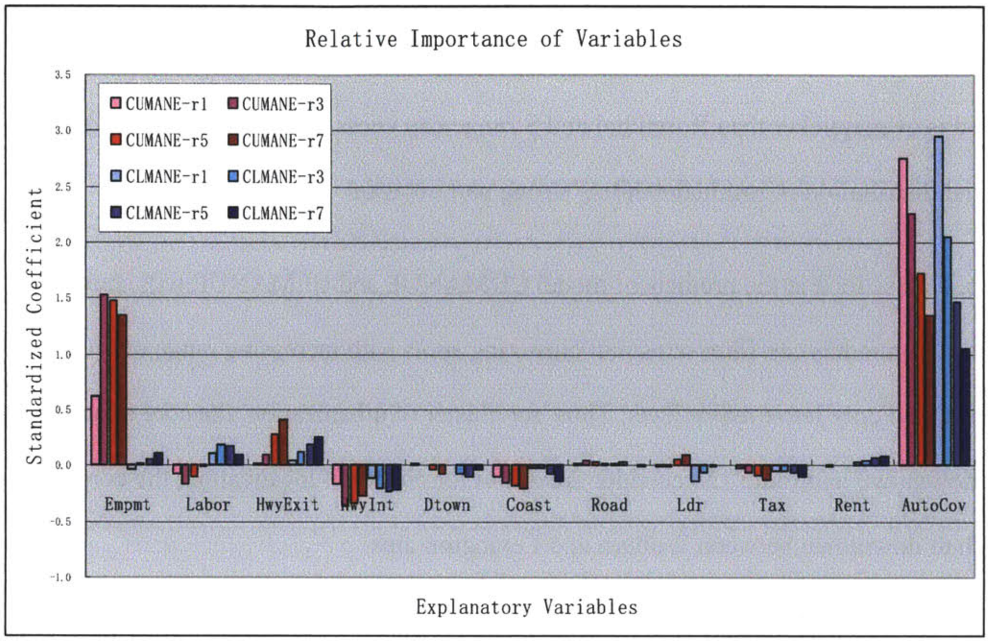
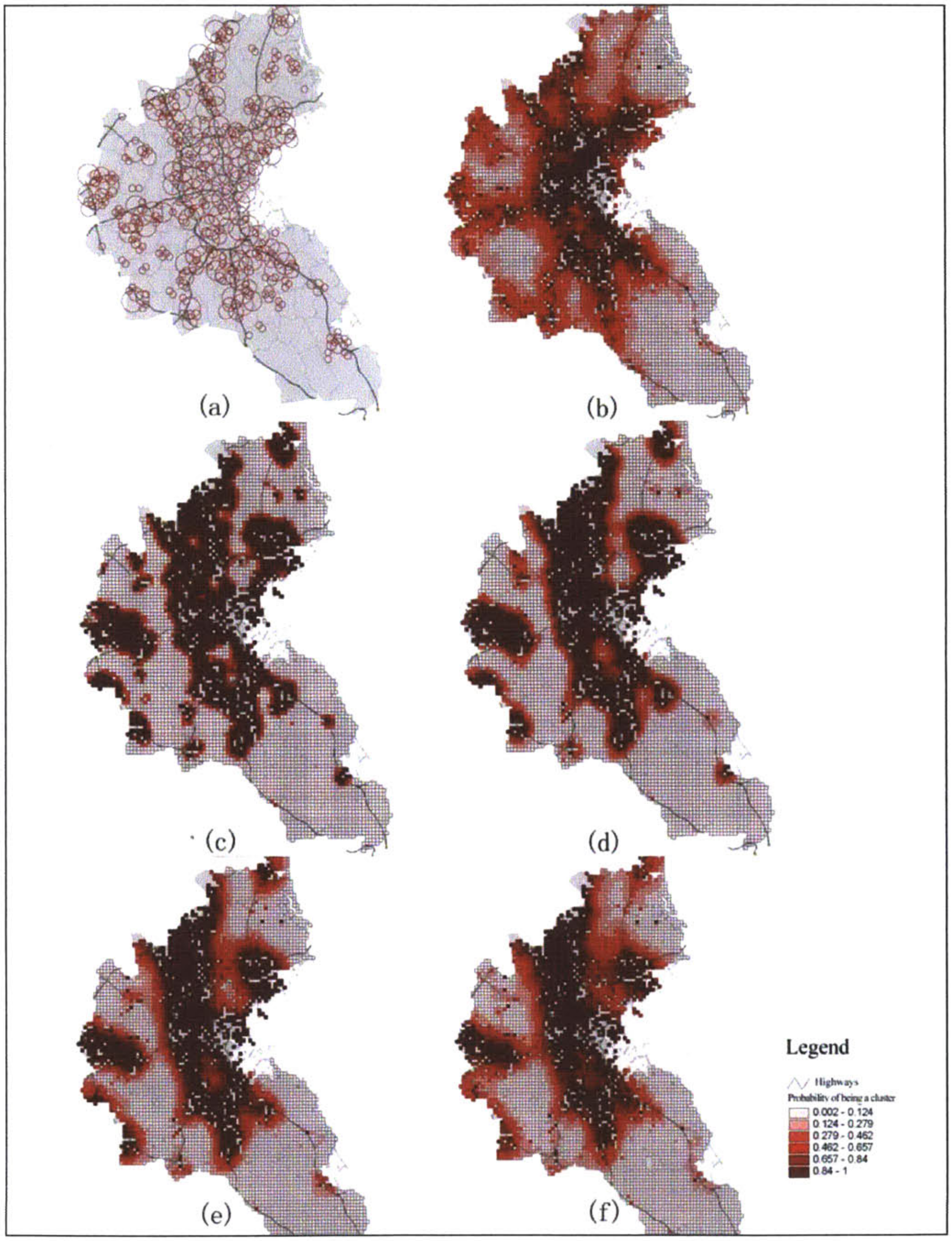


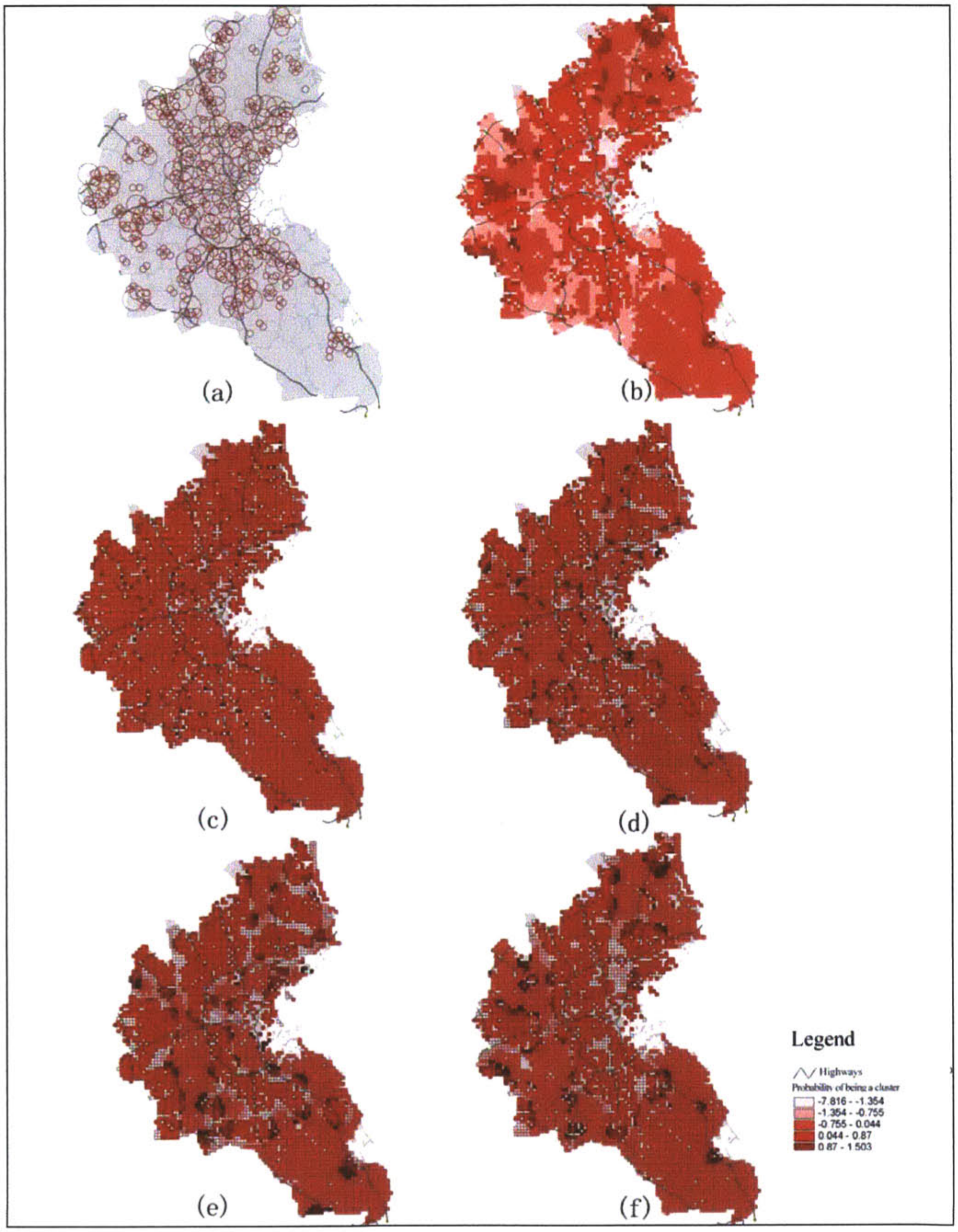
Fig 7-4 and 7-5 presents the predicted values and residuals of model CUMANUE, with spatial autocorrelation term included. From the comparisons of residuals from models with and without spatial autocovariate terms, we can see that the initial model of CUMANUE over-estimates the downtown areas, northern and western portions of the study region, but underestimate the areas between route 128 and highway 495, especially the southern portion. The non random residual distribution obviously indicates the missing of spatial relationship in the initial model. After including the spatial autocovariate term, the spatial randomness of residuals is much improved (see Fig 7.5 b, c, d, e and f), especially when small ranges of weight matrix is used. If we can logically believe that the physical concentration of firms reflects the strength of agglomeration economies, this verifies Rosenthal and Strange (2003)'s findings that agglomeration economies attenuate quickly in the first a few miles. And more likely, this attenuation speed is even quicker than Rosenthal and Strange can guess, since after 3 miles, the spatial distribution of residual exhibits strong pattern again.

Also, a close look at the predicts of model CUMANUE and CLMANUE will show that the spatial models can filter out small clustering spots with increasing range of weight matrix, and only leave major facts. Thus, for cluster of urbanization, the core areas are the downtown and route 128 corridor areas, but for cluster of localization, the core area is north to downtown between Woburn and Lexington area.


(a)
(b)
(c)
(d)
(e)
(f)
Legend
Highways
Probability of being a cluster
0.002 - 0.124
0.124 - 0.279
0.279 - 0.462
0.462 - 0.657
0.657 - 0.84
0.84 - 1

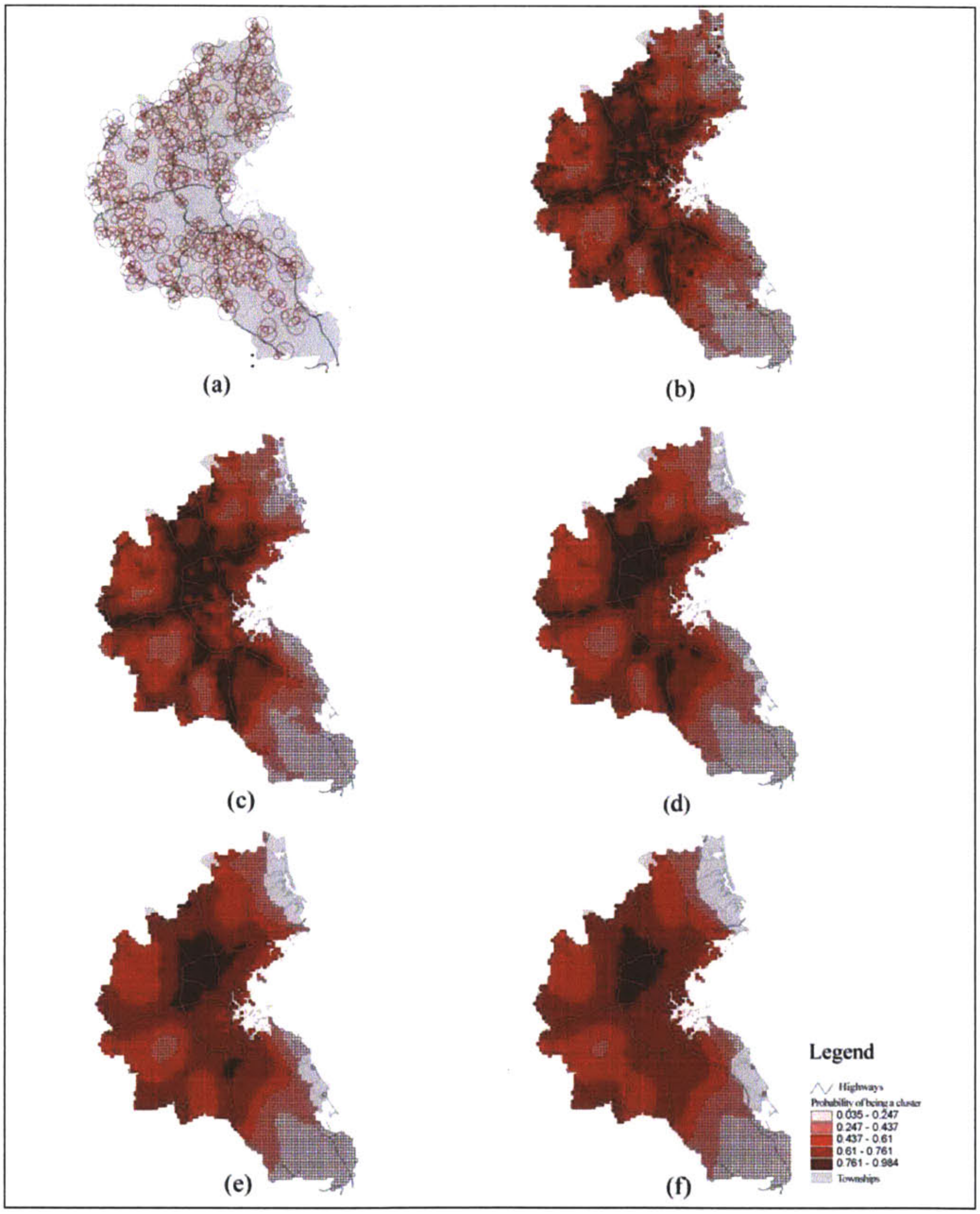
Note: (a) clusters of urbanization defined by manufacturing establishment; (b) predicted clusters of initial model; (c) predicted clusters of model radius-1; (d) predicted clusters of model radius-3; (e)predicted clusters of model radius-5; (f)predicted clusters of model radius-7;

Figure 7-4 Predicted Values of Model CUMANUE



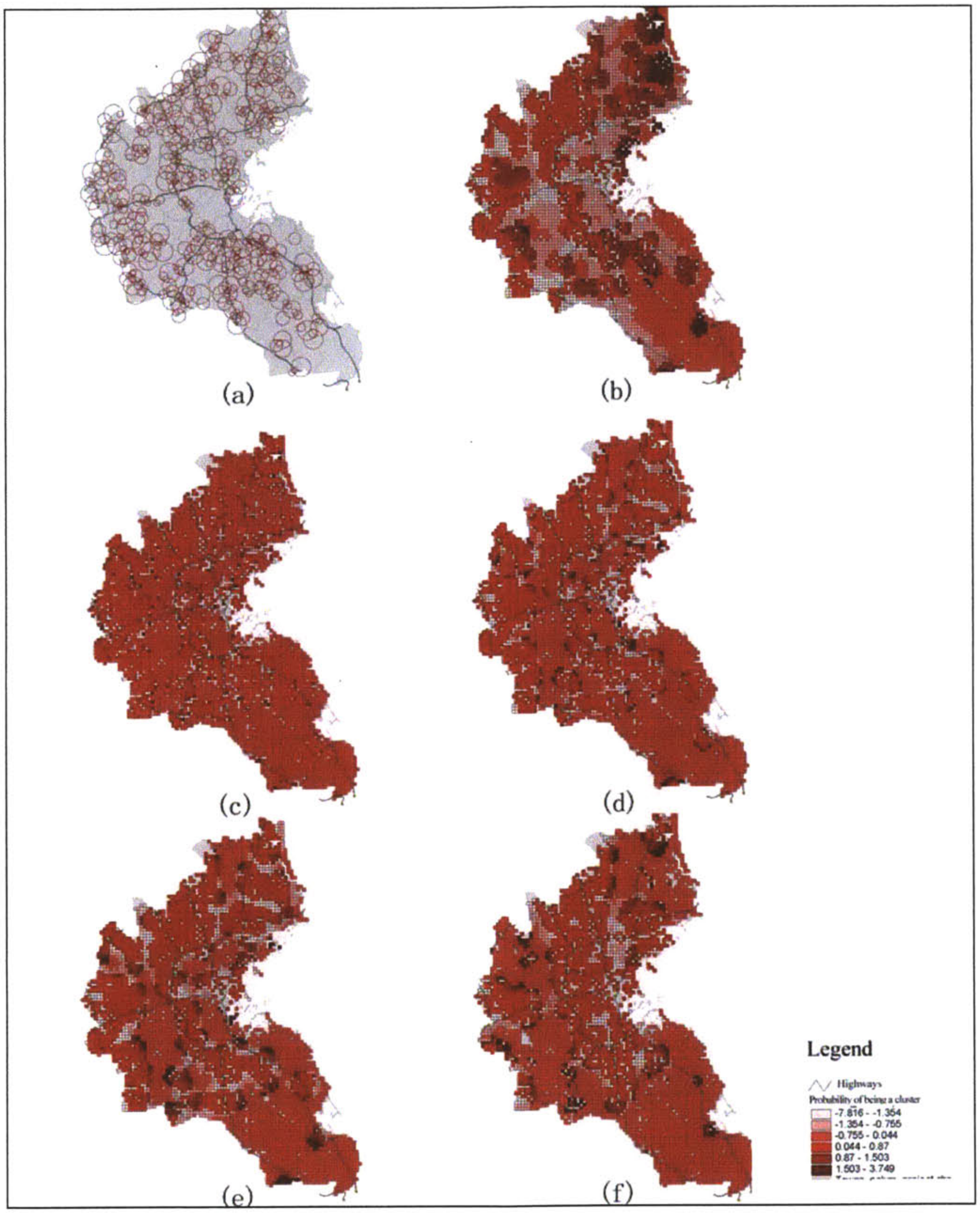
*(a) clusters of urbanization defined by manufacturing establishment; (b) residual of model CLE; (c) residual of model radius-1; (d) residual of model radius-3; (e) residual of model radius-5; (f) residual of model radius-7;

Figure 7-5 Residuals of Model CUMANUE



Note: (a) existing clusters of localization defined by manufacturing establishment; (b) predicted clusters of initial model; (c) predicted clusters of model radius-1; (d) predicted clusters of model radius-3; (e)predicted clusters of model radius-5; (f)predicted clusters of model radius-7;

Figure 7-6 Predicted Values of Model CLMANUE



*(a)existing clusters of localization defined by manufacturing establishment; (b) residual of model CLE; (c) residual of model radius-1; (d) residual of model radius-3; (e) residual of model radius-5; (f) residual of model radius-7;

Figure 7-7 Residuals of Model CLMANUE

7.6 Conclusions

In this chapter, we continue our study of the determinants of firm clustering in chapter 6 with focusing on the locational determinants of firm clusters. It is easy to understand that locational attributes affect firms' production as many of the production factors like land, labor, material inputs, tax even techniques are location bounded, but how these factors associate with firms' location decisions relative to clusters are largely left unexamined. In this chapter, with the help of detected firm clusters in chapter 4 and 5, we analyze the association between locational attributes and 4 types of clusters: urbanization cluster of establishment, urbanization cluster of employment, localization cluster of establishment and localization cluster of employment. Our results indicate that those locations with better accessibility to highway exits, downtown area and major road, lower taxes for industries, higher rental price, larger employment and bigger distance from costal lines are more likely to be identified as firm clusters. Moreover, if they are with better accessibility to highway intersections and with larger land development intensity, they are more likely to be urbanization clusters, otherwise, they are more likely to be identified as localization clusters. The unevenly distribution of firm employment size within manufacturing industry affect the magnitudes of the estimates, but clusters of establishment and employment come up with same directional findings about the association between a location's spatial attributes and its chance of being found as a firm cluster with the only exception of model CLE, which observe negative coefficient for variable labor in contrary to the positive coefficients in other models.

To avoid the prejudice caused by filtering certain overlapping clusters, we study the association between firm clustering intensity, which is the amount of overlapping clusters observed at a location, and a location's attributes. This study gives us very close findings as earlier results in this chapter, suggesting that our modeling strategy of firm clusters can give us consistent and valid findings. Moreover, the findings in the chapter verifies the findings of chapter 6 concerning that firms pay higher rental expense in locating within firm clusters. In this chapter, we also find out that in all models except CUE, we observe significant and positive coefficient for variable rent. Here we confirm that while firm clusters are defined as spatial concentration of firm establishments, they are also concentration spots of employments.

In literatures, while researchers have advocated that urbanization economy and localization economy coexist to sustain the agglomeration economies and cities, few of them notice that these two economies might cause displaced firm concentrations:. the former one creates centrifugal strength and emphasize the firm clustering at urban centers, the latter one produces centripetal forces to attract firms to locations along beltways in the suburban areas. Although our findings in this chapter and chapter 6 do not directly give reference about agglomeration economies, but the association between firm characteristics and their clustering, between firm clusters and their locational characteristics provide an hypothesis that localization economies are the main agglomeration forces within localization clusters, and similarly urbanization economies are the major driving force behind urbanization clusters. This would be a worthwhile research direction for further studies.

This chapter we also look into the spatial autocorrelation phenomenon within firm clustering. The empirical results from models show that the spatial autocorrelation significantly explaining locations' chance of falling within firm clusters. The impacts of spatial autocorrelation are the strongest at the most proximity. With the growth of distance, the strength of spatial autocorrelation sharply drops and other explaining factors start to take over.

CHAPTER 8: CONCLUSIONS AND AREAS OF FURTHER RESEARCH

8.1 Introduction

In this thesis, we attempts to construct a conceptual and computational framework for studying firm clustering at intrametropolitan level. Specifically, this framework includes methods of measuring general industry clustering, detecting sized urbanization clusters and localization clusters, interrelating firm clustering with firms' characteristics and interpreting the appearance of firm clusters at a location with its spatial settings. We apply these methods with firm establishment data to study the patterns and determinants of firm clustering within metropolitan Boston. At the first time, this study enables us to locate urbanization clusters and localization clusters in a metropolitan area and examine their association with economic and spatial explanting factors, which provides a completely new perspective of observing those spatial economical phenomenon like agglomeration and suburbanization and makes a solid step towards empirical modeling and understanding them at the micro-level. As validated in this thesis, clustering of firms or employments both end up with significant concentration of employment at some spots with specific attributes. This enables regional planners and policy-makers to apply some findings from this study in improving existing firm clustering patterns or incubating new firm concentrations, which ultimately promotes local employment and local economics.

Few studies of industrial clustering at micro level have been conducted due to the limit of disaggregated datasets and the lack of modelling method (Wu, 1999). In traditional research, industrial clustering is often treated as an elusive concept and its spatial attributes such as location, extents and social-economic surroundings are largely neglected, although industrial clustering is widely regarded as a spatial phenomenon. This study is an effort to fill this knowledge gap.

After reviewing literature in chapter 2, each chapter of this thesis is devoted to answering one of the five research questions raised at the beginning of this thesis. Chapter 3 introduces methods to measure the general industry clustering. Specifically it compares two types of indices: non-spatial indices and spatial indices and discuses their limitations; Chapter 4 introduces the method of detecting urbanization clusters; Empirically it identifies the significant urbanization clusters of establishment and urbanization clusters of employment within metropolitan Boston and compares clusters of firms of different size and technical intensity; chapter 5 continues chapter 4's work by introducing the method of detecting localization clusters and duplicates chapter 4' empirical studies of firm clustering at metropolitan Boston in terms of localization clusters. Chapter 6 introduces the modeling strategy to interpret firms' clustering decisions, specifically their site locations relative to firm clusters. This chapter empirically examines the association between firms' characteristics and their relative locations to detected firm clusters within metropolitan Boston, and investigates the joint effects of firms technical expense and productivity and also looks into the association between those relocated firms' clustering decisions and their characteristics; In chapter 7 we constructs two statistical models to examine the association between locations' spatial attributes and their chance of being a firm clustering or the local clustering intensity. In this chapter, we estimate the association magnitudes and directions between locations' attributes with detected firm clusters and clustering intensities within metropolitan Boston. We also look into the spatial autocorrelation issues within the data and identify its magnitude and attenuation patterns.

8.2 Summary of results

This thesis first measures the overall industrial clustering within Boston metropolitan area with use of two groups of indices: non-spatial indices, which include location quotient, herfindahl index, gini index and EG concentration index, and spatial indices, which includes L(r) function and D(r) function. The results of non-spatial indices identify the existence of industrial clustering in the study area. However, the EG concentration index was found to be inconsistent with other indices. The results of spatial indices further find that the level of industrial clustering in Boston Metro varies with respect to spatial scales: within the ranges of 1km-2km and 4.5km-7km, industrial clustering is most significant. The measurements of clustering in different sub-sectors such as food, machinery and instruments also find that different sectors exhibit dissimilar clustering patterns.

While the industry clustering measurements can tell us the general clustering at regional level, they can lead to incorrect conclusions when aggregating clustering and dilution of firms at regional level. Also, industry clustering measurements are unable to identify the exact extent and location of firm clusters, which largely limits our further understanding of firm clustering at micro level. Our subsequent effort was to detect clusters by identifying industrial clusters' locations and extents, and statistically evaluate their significance. The results show ubiquitous existence of overlapping firm clusters in all spatial scales within metropolitan Boston. Also, clusters of firms of different sizes and knowledge intensities exhibit diverse spatial patterns.

These findings demonstrate that the distribution of firm clustering composes of a complex system of clusters with various locations and extents. The clustering is a relative concept that exists upon its comparison reference, which is often the study area. Conclusions about firm clustering without reference to its applied scale and its study area usually creates confusions unless they are validated by case studies of all situations.

The study also shows that urbanization clusters and localization clusters are with very dissimilar distributions. Furthermore, clusters of establishment and clusters of employment are also different due to the unevenly distribution of firm size within manufacturing industry. These concepts do not divide the firm clusters into exclusive subcategories, but they do differentiate firm clusters at several spatially divergent groups.

Furthermore, using detected firm clusters, this study attempts to investigate how and why firms demonstrate different clustering behaviors such as their location relative to detected clusters. We then analyzed the association between firms' characteristics and their locations relative to firm clusters. The empirical results strongly suggest that firms within clusters are significantly different from those outside of clusters. Firms with larger employment size, productivity and rent expense are more likely to be found in either urbanization clusters or localization clusters. Attributes like productivity, technical expense, advertise expense, utility expense and branch or not also associate with firms' location relative to clusters, but the exact magnitudes and direction depends on which types of clusters is discussed. For firms in localization clusters, firms with higher technical expense and higher productivity cluster together, suggesting the existence of active R&D activities and localization economies. Firms within localization clusters are also found paying less utility expense, which can be the result of infrastructure sharing among many firms in the same industry. In urbanization clusters, firms are found with higher utility cost and insignificant productivity advantages compared with firms outside of clusters. Branch firms are more likely to be found within clusters, except large branch firms, which are found negatively associated with localization clusters.

The study of firms relocated between 2004 and 2009 indicate that firm clustering is a dynamic process, instead of a static snapshot. Firms adjust their clustering behaviors according to their new demands. For example, we found those relocated firms with larger employee size are more likely to be found within clusters, but the smaller firms are more likely to be found outside. This process ends up with large firms' concentration within firm clusters, as validated by empirical results.

Finally, we investigate the locational determinants of firm clusters. The empirical results indicate that those locations with better accessibility to highway exits, downtown area and major road, lower taxes for industries, higher rental price, larger employment and bigger distance from costal lines are more likely to be found as firm clusters. Furthermore, if they are with better accessibility to highway intersections and with larger land development intensity, they are more likely to be urbanization clusters, otherwise, they are more likely to be identified as localization clusters. The unevenly distribution of firm employment size within manufacturing industry affect the magnitudes of the estimates, but clusters of establishment and employment come up with same directional findings about the association between a location's spatial attributes and its chance of being found as a firm cluster with the only exception of model CLE, which observe negative coefficient for variable labor in contrary to the positive coefficients in other models.

To avoid the prejudice caused by filtering certain overlapping clusters, we study the association between firm clustering intensity, which is the amount of overlapping clusters observed at a location, and a location's spatial attributes. The results show very close findings to what we find with firm cluster models, suggesting that our modeling strategy of firm clusters can give us consistent and valid findings.

In literatures, while researchers have advocated that urbanization economy and localization economy coexist to sustain the agglomeration economies and cities, few of them notice that these two economies might cause displaced firm concentrations:. While the former one creates centrifugal strength and emphasizes the firm clustering at urban centers, the latter one produces centripetal forces to attract firms to locations along beltways in the suburban areas. Although our findings in this thesis do not directly give reference about agglomeration economies, the association between firm characteristics and their clustering, between firm clusters and their locational characteristics suggest an hypothesis that localization economies are the main agglomeration forces within localization clusters, and similarly urbanization economies are the major driving force behind urbanization clusters.

The empirical results also verify that the spatial autocorrelation significantly exist within the data. The impacts of spatial autocorrelation are the strongest at the most proximity. With the growth of distance, the strength of spatial autocorrelation sharply drops and other explaining factors of firm clustering start to take over.

8.3 Policy Implications

Firm clustering is adopted by more and more local governments as a development tool to promote local economic and employment growth. Although the facts about firm clustering discovered within this empirical study are far from enough to set up instructional theories, they do provide several important policy implications for firm clustering related policy practice. First, as firm clustering exhibit different patterns and are driven by dissimilar dynamics at different spatial scales, the cluster development policies that clearly define the targeting scale and treat each scale differently will have better effects in creating planned clusters; second, as discovered in this study, urbanization clusters and localization clusters exhibit big differences in location preferences and composition, policy makers need choose the appropriate type of clusters that best fit in with local conditions; third, this study also finds that firms response differently toward clustering due to their characteristics differences, thus after choosing the targeting spatial scale and the type of firm clusters, policy makers also need specify the firms that they plan to attract. Targeting those firms that prefer to clusters will obviously promote the policy effects; finally, as not every place is identically attractive for firm clusters, policy makers also need carefully choose appropriate locations according to the targeting spatial scale, type of clusters and firms.

8.4 Further studies

The research in this thesis presents significant and interesting results about firm clustering, but limited by research time and funding, it also poses more questions and lines for further levels of research. Three central topics can be studied in more detail to get a better understanding of firm clustering empirically: the agglomeration economies within each type of firm clusters, a time series analysis of the firm clustering process and an integrated model for understating the interactions between firms and locations during firm clustering process.

REFERENCE

Amin, A. and N. Thrift. 1994. " Living in the Global", in A. Amin and N. Thrift (eds.). Globalization, Institutions and Regional Development in Europe. Oxford: Oxford University Press.

Amiti, Mary (1999), Specialization Patterns in Europe, Weeltwirtschaftliches Archiv, 134 (4), 573-593

Anand, Sudhir, 1983, *Inequality and Poverty in Malaysia: Measurement and Decomposition*, Oxford University Press, London. (Anand 1983)

Appold, S. J. 1995. "Agglomeration, Interorganizational Networks, and Competitive Pedormance in the U.S. Metalworking". *Ecoiiomic Geograplty*. 71(1): 27-54

Atkinson, A. B. and Bourguignon, F. (2000). Introduction: Income Distribution and Economics, in A. B. Atkinson and F. Bourguignon (eds.) Handbook of Income Distribution: Volume 1. Amsterdam: Elsevier.

Besag J E, 1977. Comments on Ripley's paper. Journal of the Royal Statistical Society, B29: 193–195

Callejón, M. (1997), 'Concentratión geográphica de la industria y economiasde aglomeratión', Economia Industrial 5, 61—68.

Chakravarty, S. R. (1990). Ethical Social Index Numbers, New York:Springer-Verlag

Devereux, M., Griffith, R. and Simpson, H. (1999), 'The Geographic Distribution of productive Activity in the UK', Institute for Fiscal Studies Working Paper, W99/26.

Diggle P J and Chetwynd A G, 1991. Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics, 47: 1155–1163

Bania, Neil; Eberts, Randall W; and Fogarty, Michael S. 1993. "Universities and the Startup of New Companies: Can We Generalize from Route 128 and Silicon Valley?" *The Review of Economics and Statistics,* MIT Press, vol. 75(4), pages 761-66, November.

Behavior Center for Urban and Regional Studies, The University of Bermingham.

Blaug, Mrk, 1997, Economic Theory in Retrospect, 5th edition. Cambridge Univeristy Press.

Briant, Anthony, Pierre-Philippe Combes, and Miren Lafourcade. 2007. Does the size and hape of geographical units jeopardize economic geography estimations? Processed, Paris-jourdan Sciences Economiques.

Brown, J.S. and Duguid, P. (2000) "Mysteries of the region: knowledge dynamics in Silicon Valley " in C-M. Lee, W.F. Miller, M.G. Hancock, and H.S. Rowen (eds) *The Silicon Valley Edge* , Stanford, CA: Stanford Business School Press, pp. 16-39.

Brülhart, Marius and Torstensson , J. 1996. Regional Integration, *Scale Economies and Industry Location,* CEPR Discussion Paper No. 1435

Carroll, Thomas M., and Dean, Robert D. 1976. "Plant Location, Search Costs and the Econmics of Information." Presented at the Western Regional Science Association meetings, San Diego, California, Feb. 25

Christaller, W. (1966 (1933)) *Central Places in Southern Germany.* Translated by C.W. Baskin. Englewood Cliffs: Prentice-Hall.

Chung, Wilbur and Kalnins, Arturs. 2001. "Agglomeration Effects and Performance: A Test of the Texas Lodging Industry." Strategic Management Journal, 22(10): 969-997.

Devereux, M.P., Griffith, R. and Simpson, H. (1999), "The Geographic Distribution of Production Activity in the United Kingdom ", The Institute for Fiscal Studies, Working Paper No. 26.

DeYoung, Robert, William C. Hunter and Gregory F. Udell, 2004, "The Past, Present and Probable Future for Community Banks," Journal of Financial Services Research25:2-3 (April/June), 85-134.

Doeringer, P and DG. Terkla. (1995). Business Strategy and Cross—industry cluster [J]. Economic Development Ouarterly, vol.9, p225.

Ellison, G., and E. L. Glaeser. 1997. Geographic concentration in U.S. manufacturing industries:A dartboard approach. *Journal of Political Economy* 105: 889-927.

Ellison, Glenn; Glaeser, Edward L. and Kerrr, William R.. 2007. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," Working Papers 07-064, Harvard Business School.

ENGLANDER, (1926) Kritisches and Positives zu einer allgemeinen reinen Lehre vom Standort, Zeitschriftftir Volkswirtschaft und Sozialpolitik, Neue Folge, 5.

Feser, E. J. 1998. "Old and new theories of industry clusters", in *Clusters and Regiona Specialisation*, edited by M. Steiner, pp. 18-40, London, Pion Limited.

Fujita M., and Krugman P. [1995]. "When in the Economy Monocentric? von Thunen andChamberlinUnified.". Regional Science and Urban Economicsn. 25, pp. 505-528.

Glaeser E, 1999. "Urban Primacy and Politics," Harvard Institute of Economic Research Working Papers 1874, Harvard - Institute of Economic Research.

Gini, Corrado (1912). Variabilit`a e Mutabilit`a, Bologna: Tipografia diPaolo Cuppini.

Gini, Corrado (1914). "Sulla Misura della Concertrazione e della Variabilit` a dei Caratteri," Atti del R. Istituto Veneto di SS.LL.AA., 73, 1203–1248.

Gini, Corrado (1921). "Measurement of Inequality of Incomes," Economic Journal, 31, 124–126.

Goldstein GS and Gronberg TJ (1984) Economies of Scope and Economies of Agglomeration. Journal of Urban Economics 16: 91104

Greenhut, Melvin. 1958. Plant Location in Theory and Practice. Chapter Hill: University of North Carolina Press

Harrison, Bennett and Amy Glasmeier, (1997). "Response: Why Business Alone Won't Redevelop the Inner City: A Friendly Critique of Michael Porter's Approach to Urban Revitalization," *Economic Development Quarterly*, Vol. 11 (1) February, 28-38.

Hamilton, F.E.I. (ed.). Spatial Perspectives on Industrial Organization and Decision-Making, London: John Wiley, 1974

Hansen, Robert G., 1987, Atheory for the choice of exchange medium in the market for corporate control, *Journal of Business* 60, 75-95

Helsley, R. and Strange, W. 1990. Agglomeration economies and matching in a ystem of cities. *Regional Science and Urban Economics* 20, 189–212.

Honjo, Y and O. Kusakabe (2000): Proposal of comprehensive foundation design code: Geo-code 21, Proc. LSD2000: International Workshop on Limit State design in Geotechnical Engineering, pp. 75-84.

Hong, J 2007. Firm Heterogeneity and Location Choices: Evidence from ForeignManufacturing Investments in China. *Urban Studies.*, September 2011 48(12)

Hope, A. C. A. 1968. A simplified Monte Carlo significance test procedure. *J. R. Stat. Soc. Ser.* B 30: 582–598.

Houdebine, M. 1999. Concentration Géographique des Activités et Spécialisation des Départements Français, *Économie et Statistique,* n°326-327, pp. 189-204.

Isard W. 1949 'The General Theory of Location and Space-Economy', *Quarterly Journal of Economics* LXIII(4):476506.

Isard, W. and T. A. Reiner. 1966. Regional Science: Retrospect and Prospect. *Papers, Regional Science Association*, European Congress, Cracow (1965)16: 1-16.

Krugman, Paul, 1980. "Scale Economies, Product Differentiation, and the Pattern of Trade," American Economic Review, American Economic Association, vol. 70(5), pages 950-59, December.

Krugman, Paul, 1991a. Geography and trade. Leuven: Leuven University Press.

Krugman,Paul, 1991b. Increasing returns and economic geography. Journal of Political

Krugman, P. 1995. *Development, Geography, and Economic Theory.* Cambridge and London: The MIT Press

Krugman, Paul and Raul Livas. 1996. "Trade Policy and the Third World Metropolis." *Journal of Development Economics* 49(1): 137-50

Krugman P. 1998. "What's new about the new economic geography?" , *Oxf. Rev. Econ. Pol.* 14:2, pp. 7-17

Krugman, P. 1998. Space: The final frontier. Journal of Economic Perspectives 12(2): 161-74.

Kulldorff, M. 1997. "A spatial scan statistic." *Communications in Statistics: Theory and Methods.* 26:1481-1496.

Lall, S. and Mengistae, T. "Business Environment, Clustering, and Industry Location: *Evidence from Indian Cities,*" World Bank Policy Research Working Paper, 2005b.

Lambert, Peter J. (1989). The Distribution and Redistribution of Income: A Mathematical Analysis, Cambridge, Massachusetts: Basil Blackwell Inc..

LeVeen, Jessica (1998) Industrial Cluster Literature Review, The University of North Carolina at Chapel Hill

Lösch, A. 1940. The Economics of Location, Jena, Germany: Fischer. (English translation, New Haven: Yale U. Press, 1954.

Logan, M.I. 1966. "Locational Behavior of Manufacturing Firms in Urban Areas." *Annals of the Association of A merican Geographers* 56:451-466.

Lösch, A. 1940. The Economics of Location, Jena, Germany: Fischer. (English translation, New Haven: Yale U. Press, 1954.

Lundquist, Karl-Johan and Lars-Olof Olander (1998), Regional economies: a threat to the nation-state? (CFE Working Paper, 3. Lund: Centre for European Studies, Lund University).

Marshall, Alfred. 1920. Principles of Economics. London: MacMillan

Martin, Ron and Sunley, Peter (1996) "Paul Krugman's Geographical Economics and its implications for regional development theory: a critical assessment", Economic Geography, Vol. 72, No3., pp. 259-292

Martin, R, 1999, The New "Geographical Turn" in Economics: Some Critical Reflections, *Cambridge Journal of Economics,* 23, pp. 63-91.

Marcon, E., & Puech, F. (2003). Measures of Geographic Concentration of Industries: Improving Distance Based Methods. Mimeo, TEAM University of Paris

Maurel F and Sedillot B, 1999. A measure of the geographic concentration in French manufacturing industries. Regional Science and Urban Economics, 29: 575–604

Meardon, Stephen J. 2001. "Modeling Agglomeration and Dispersion in City and Country: Gunnar Myrdal, Francois Perrous, and the New Economic Geography". *American Journal of Economics and Sociology*. Vol. 60. No11. Special Issue: City and Country: An Interdescipplinary Collection (Jan., 2001), pp. 25-37

Midelfart-Knarvik ,Karen; Overman,Henry and Venables, Anthony. 2001. Comparative Advantage and Economic Geography, CEPR Discussion Paper No 2618, draft September 2001

Miller, Mark S., Lay Gibson, and Gene Wright. 1991. "Location Quotient: A Basic Tool for Economic Development Analysis." *Economic Development Review* 9(2): 65-68.

Ohlin, Bertil. 1933. Interregional and International Trade. Cambridge, MA: HarvardUniversity Press.

Okamuro, H., 2008, How different are the regional factors of high-tech and low-tech start-ups? Evidence from Japan, International Entrepreneurship and Management Journal 4, 199-215.

Openshaw, S., M. Charlton, C. Wymer, and A. Craft (1987). "A Mark 1 Geographical Analysis Machine for the automated analysis of point data sets." *International Journal of Geographic Information Systems*. 1:335-358

Ottaviano, Gianmarco and Puga, Diego (1997). "Agglomeration in the Global Economy: A Survey of the New Economic Geography". CEPR Discussion Paper. Nº 1.699. Londres.

Peroux, F. 1950. "Economic Space: Theory and Application." *Quarterly Journal of Economics.*

Phelps, N., Fallon, R. and Williams, C. (2001), "Small Firms, Borrowed Size and the Rural-Urban Shift", *Regional Studies*, vol. 35, pp. 613-624.

Porter, Michael E. (1997). New strategies for inner-city economic development. *Economic Development Quarterly*, 11(1):11-27.

Porter, M. E. 1998. *The Competitive Advantage of Nations (with a new foreword)*, New York: The Free Press.

Pred, A.R. 1964 'The intrametropolitan location of American manufacturing' *Annals of the Association of American Geographers* 54, 165– 180.

Predöhl, Andreas.1928. The theory of Location in its Relation to Genral Economics. Journal of Political Economy. XXXXVI, 371-90

Puga, Diego & Venables, Anthony J., 1996. "The Spread of Industry: Spatial Agglomeration in Economic Development," CEPR Discussion Papers 1354, C.E.P.R. Discussion Papers.

Puga, Diego & Venables, Anthony J., 1997. "Trading arrangements and industrial development," Policy Research Working Paper Series 1787, The World Bank

ReVelle, C., & Laporte, G. 1996. "The Plant Location Problem". *New Models and Research Prospects OR Chronicle, 44*(6), 864-874.

Rosenthal, Stuart S. and William C. Strange, "The Determinants of Agglomeration, Journal of Urban Economics 50(2), 2001, 191-229.

Sabel, C. (1989). "Flexible Specialization and the Re-emergence of Regional Economies" in P. Hirst and J. Zeitlin (eds.) Reversing Industrial Decline, New York: Berg. Ch. 1.

Saxenian, AnnaLee. Regional advantage: culture and competition in Silicon Valley and Route 128. Cambridge, MA: Harvard University Press, 1994

Schmutzler, A. (1999) The new economic geography, Journal of Economic Surveys, 13, 355 – 79.

Scott, Allen J.1998. Regions and the World Economy, The coming shape of Global Production,Competition and political Order. paperback edition 2000 ed. Oxford, New York: Oxford University Press.

Silber, Jacques (1999). Handbook of Income Inequality Measurement, Boston : Kluwer Academic.

Storper, M. 1993. Regional "worlds" of production: Learning and innovation in technology districts of France, *Italy and the USA. Regional Studies*, 27: 433–456.

Struyk R, James F. 1975 Intrametropolitan Industrial Location: The Patterns and Process of Change .D.C. Heath, Lexington, MA

Swann, G.M.P. and Prevezer, M. (1996), A comparison of the dynamics of industrial clustering in computing and biotechnology, Research Policy 25: 1139-1157.

Sweeney, S. H., and E. J. Feser. 1998. Plant size and clustering of manufacturing activity. *Geographical Analysis* 30: 45-64.

Townroe, P.M. 1971. Industrial Location Decisions. A study in Management

Terkla, D and Doeringer, P. (1995). Business strategy and cross-industry clusters. *Economic Development Quarterly*, 9, 225-37.

Valeyre, A. 1993. Mesures de dissemblances et d'inégalité interrégionales: principes, formes et propriétés, *Revue d'Economie Régionale et Urbaine* 1: 17-53

Valeyre, A. (1993) Measures de dissemblance et d'inegalite interregionales: principes, formes et proprietes. Revue d'Economie Regionale et Urbaine, 1:17-53.

VEDP/VCCS Cluster Study, 2005, Virginia Economic Development Partnership (VEDP)/Virginia Community College System (VCCS), http://www.virginiaallies.org/research.asp

Venables, A.J., 1996. Equilibrium locations of vertically linked industries. International Eco-nomic Review 37(2), 341–358.

Von Thunen, J . 1826. The isolated state, English edn. London: Pergamon Press.

Webber, A. 1929. [Translated by C.J. Friedrich] Theory of the Location of Industries: The University of Chicago Press: University Press:New Haven, CT, USA

Webber, Michael. 1972. Impact of Uncertainty on Location. Cambridge: MIT Press

Wiig, H. & Wood, M. (1995), What Comprises a Regional Innovation System? – An Empirical Study, STEP Working Paper R-01, Oslo

Wolff, James A. and Timothy L. Pett (2000). "Internationalization of small firms: An examination of export competitive patterns, firms size, and export performance." Journal of Small Business Management. Vol. 38, No. 2, pp. 34 – 47.s

Wu, F. 1999. "Intrametropolitan FDI Firm Location in Guangzhou", China. *The Annals of Regional Science* 33, pp. 535–555